NBER WORKING PAPER SERIES

DEMAND ANALYSIS USING STRATEGIC REPORTS: AN APPLICATION TO A SCHOOL CHOICE MECHANISM

Nikhil Agarwal
Paulo Somaini

Working Paper 20775
http://www.nber.org/papers/w20775

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2014

We are grateful to Parag Pathak for sharing data on admissions in Cambridge Public Schools, and for several helpful discussions. We thank Steve Berry, Phil Haile and Mike Whinston for several helpful discussions, and Susan Athey, Lanier Benkard, Victor Chernozhukov, William Diamond, Liran Einav, Whitney Newey, Ariel Pakes, Larry Samuelson, Joe Shapiro and Frank Wolak for their comments. Vivek Bhattacharya provided outstanding research assistance. Both authors thank the Cowles Foundation of Economic Research and the Department of Economics at Yale University for their generous hospitality. Support from the National Science Foundation (SES-1427231) is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Demand Analysis using Strategic Reports: An application to a school choice mechanism
Nikhil Agarwal and Paulo Somaini
NBER Working Paper No. 20775
December 2014
JEL No. C50,D47,I20

ABSTRACT

Several school districts use assignment systems in which students have a strategic incentive to misrepresent their preferences. Indeed, we find evidence suggesting that reported preferences in Cambridge, MA respond to these incentives. Such strategizing can complicate the analysis of preferences. The paper develops a new method for estimating preferences in such environments. Our approach views the report made by a student as a choice of a probability distribution over assignment to various schools. We introduce a large class of mechanisms for which consistent estimation is feasible. We then study identification of a latent utility preference model under the assumption that agents play a Bayesian Nash Equilibrium. Preferences are non-parametrically identified under either sufficient variation in choice environments or sufficient variation in a special regressor. We then propose a tractable estimation procedure for a parametric model based on Gibbs' sampling. Estimates from Cambridge suggest that while 84% of students are assigned to their stated first choice, only 75% are assigned to their true first choice. The difference occurs because students avoid ranking competitive schools in favor of less competitive schools. Although the Cambridge mechanism is manipulable, we estimate that welfare for the average student would be lower in the popular Deferred Acceptance mechanism.

Nikhil Agarwal
MIT
Department of Economics, E17-224
77 Massachusetts Ave
Cambridge, MA 02139
and NBER
agarwaln@mit.edu

Paulo Somaini
Department of Economics, E18-212
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
psomaini@mit.edu

A data appendix is available at:
http://www.nber.org/data-appendix/w20775

1 Introduction

Admissions to public schools in the United States and abroad commonly use assignment mechanisms based on student priorities, a lottery, and importantly, reported ranking of various school options (Abdulkadiroglu and Sonmez, 2003; Pathak and Sonmez, 2008). Data on reported rankings generated by these mechanisms promise several opportunities for academic research and directing school reforms. However, with rare exceptions, mechanisms used in the real world are susceptible to gaming (Pathak and Sonmez, 2008), making it difficult to directly interpret reported lists in as true preference orderings. Table 1 presents a partial list of mechanisms in use at school districts around the world. To our knowledge, only Boston currently employs a mechanism that is not manipulable.[1]

[Table 1 about here.]

Previous empirical work has typically assumed that observed rank order lists are truthful representation of the students' preferences (Hastings et al., 2009; Abdulkadiroglu et al., 2014; Ayaji, 2013), allowing a direct extension of discrete choice demand methods with such data (c.f. McFadden, 1973; Beggs et al., 1981; Berry et al., 1995, 2004).[2] The assumption is usually motivated by properties of the mechanism or by arguing that strategic behavior may be limited under a sudden change in the choice environment. This standard approach may not be valid if students have a strategic incentives to manipulate their reports. Anecdotal evidence from Boston suggests that parent groups and forums for exchanging information about the competitiveness of various schools and discussing ranking strategies are fairly active (Pathak and Sonmez, 2008). Laboratory experiments also suggest that agents participating in manipulable mechanisms are more likely to engage in strategic behavior (Chen and Sonmez, 2006; Calsamiglia et al., 2010).

This paper proposes a general method for estimating the underlying student preferences for schools using data from manipulable mechanisms. We make several methodological and empirical contributions. Our main empirical contributions are an analysis of strategic behavior in elementary school admissions Cambridge, an application of our methods to estimate preferences, and counterfactual analysis of an alternative school choice mechanism. Our main technical contributions include a new class of mechanisms for which preference

[1]The student proposing deferred acceptance mechanism is a commonly used mechanism that is strategy-proof if students are not restricted to list fewer schools than are available. However, with the exception of Boston since 2005, all implementations of the mechanism known to us, severely restrict the length of the rank-order list. Abdulkadiroglu et al. (2009) and Haeringer and Klijn (2009) show that with this restriction, the mechanism provides incentives for students to drop competitive schools from their rank-order list.

[2]He (2012) and Calsamiglia et al. (2014) are notable exceptions that allows for agents to be strategic. We compare our results with this paper in further detail below.

parameters can be consistently estimated, a study non-parametric identification in such an environment, and a computationally tractable estimator. These innovations allow for using our methods to estimate preferences in manipulable school choice mechanism, while accounting for strategic behavior.

Ignoring the possibility of strategic behavior can have important implications. First, school accountability and improvement programs, or district-wide reforms, are liable to using stated rank order lists as direct indicators of school desirability or student preferences. For instance, Boston's Controlled Choice Plan used the number of applications to a school as a formal indicator of school performance in a school improvement program. Several manipulable mechanisms provide students incentives to avoid reporting competitive schools, and therefore, using stated rank lists reported to inform policy can misdirect resources. Second, recent empirical studies in economics have used estimates of student preferences assuming truthful behavior to evaluate student welfare under alternative matching mechanisms (see Abdulkadiroglu et al., 2014, for example). Strategyproof mechanisms are advocated on the basis of their simplicity, robustness to information available to participants and fairness (see Azevedo and Budish, 2013, and references therein). While it is well-known that such mechanisms may compromise student welfare (Miralles, 2009; Abdulkadiroglu et al., 2011), we are able to quantify potential welfare costs of switching to such a mechanism. Third, recent studies have used preference estimates for studying implications for student achievement (Hastings et al., 2009), and school competition (Nielson, 2013). These approaches may not be suitable for data from manipulable assignment mechanisms if strategic behavior is widespread.

Our analysis of ranking behavior for admissions into public elementary schools in Cambridge indicates significant gaming. The school district uses a variant of the Boston Mechanism, that is highly manipulable. We find large strategic incentives in this school system: some schools are rarely assigned to students that rank it second, while others are have spare capacity after all students have been considered. Students therefore risk losing their priority if they do not rank one of the competitive schools as their first choice. We investigate whether students appear to respond to these incentives using a regression discontinuity design. The design leverages the fact that students receive proximity priority at the two closest schools. We find that student ranking behavior changes discontinuously with the change in priority. This finding is not consistent with a model in which students state their true preferences, and a distribution of preferences that is continuous with respect to distance.

Therefore, instead of interpreting stated rank order lists as true preferences, our empirical approach is based on interpreting a student's choice of a report as a choice of a probability distribution over assignments. Each rank-order list results in a probability of getting assigned to each of the schools on that list. This probability depends on the student's priority type and report, a randomly generated lottery (if there is one), as well as the reports and priorities of the other students. If agents have correct beliefs about this probability and are expected utility maximizers, the chosen report then reveals comparisons of expected utilities with other reports the agent could have chosen. Formally, we require that student behavior is described by a Bayesian Nash Equilibrium. This assumption implies sophisticated agent behavior an is an important baseline model for accounting the strategic behavior observed.

In order to learn about preferences from the reports made by a student, we first estimate the probabilities of assignment associated with each report and priority type. Constructing consistent estimates of these probabilities requires a consideration of potential dependent data since the assignment of an individual agent depends on the reports of all other agents in the economy. We present a general convergence condition on the mechanism under which data from a large market can be used to consistently estimate these probabilities without directly estimating preferences or solving for an equilibrium. The ability to do this circumvents difficulties that may arise due to computational difficulties in solving for an equilibrium or multiplicity of equilibria.

A priori, this convergence condition can be hard to verify because assignment mechanisms are usually described in terms of algorithms rather than functions with well-known properties such as continuity. We therefore introduce a new class of mechanisms called report-specific priority + cutoff mechanisms for which we prove that this condition is satisfied. All mechanisms in Table 1, except the Top Trading Cycles mechanism, can be represented as report-specific priority + cutoff mechanisms. Our results additionally require that a lottery is used to break ties in assignments and that there are coarsely defined priority types. This rules out admissions in some school districts for exam schools or other programs that use test scores to determine admissions.

Since the assignment probabilities as a function of reports and priority types can be consistently estimated, we study identification of preferences treating these probabilities as known both to the econometrician and to agents that maximize expected utility. The problem is equivalent to identifying the distribution over preferences over discrete objects with choice data on lotteries over these objects. Indeed, the classical discrete choice demand model is a special case with degenerate lotteries. We follow the discrete choice literature in specifying preferences using a flexible random utility model that allows for student and school unobservables (see Block and Marshak, 1960; McFadden, 1973; Manski, 1977). We show conditions under which the distribution of preferences is non-parametrically identified.

We exploit two types of variation to identify the distribution of preferences. First, we use variation in choice environments (as defined by the lotteries available to the agents). Such variation may arise from differences in agent priorities that are excludable from preferences, or if the researcher observed data from two identical populations of agents facing different mechanisms or availability of seats. We characterize the identified set of preference distributions under such variation. Although sufficient variation in choice environments can point identify the preference distribution, we should typically expect set identification. Our second set of identification results relies on the availability of a special regressor that is additively separable in the indirect utility function (Lewbel, 2000). The assumption is commonly made to identify preferences in discrete choice models (Berry and Haile, 2010, 2014, for example). We show that local variation in this regressor can be used to identify the density of distribution of utility in a corresponding region. A special regressor with full support can be used to identify the full distribution of preferences.

We propose an estimation procedure for the distribution of preferences using a Gibbs' sampler adapted from McCulloch and Rossi (1994).[3] The estimator lends itself naturally to our setting because the set of utility vectors for which a given report is optimal can be expressed as a convex cone. This allows us to implement an estimation procedure that does not involve solving for the optimal report given a simulated draw of the utility.

We apply this two-step method to estimate student preferences in Cambridge. The estimated preferences can be used to address a wide range of issues. We investigate the extent to which students avoid ranking competitive schools in order to increase their chances of assignment at less competitive options. Prevalence of such behavior can result in certain schools mis-estimating the attractiveness of schools if stated ranks were interpreted on face value. Ignoring strategic behavior may therefore result in inefficient allocation of public resources to improve school quality. Further, a large number of students assigned to their first choice may not be an indication of student satisfaction or heterogeneity in preferences. We therefore investigate if strategic behavior results in fewer students are assigned to their true first choice as compared to their stated first choice.

Finally, we compare the welfare effects of a switch to the student proposing Deferred Acceptance Mechanism. The welfare comparison between the two mechanisms is theoretically ambigious. The mechanisms also cannot be empirically evaluated without estimating preferences because of strategic reporting. Our method recovers the true preference distribution when agents are strategic, which allows us to compare welfare effects of counterfactual designs of the market.

Related Literature

[3] We view our non-parametric identification results as justifying that parametric assumptions are not essential for learning about the primitives of interest but are made to assist estimation in finite samples.

Our approach to studying large sample properties of our estimator and defining a limit mechanism is motivated by recent theoretical work studying matching markets by Kojima and Pathak (2009) and Azevedo and Leshno (2013). Some of our proposed results rely on and extend the large market results in Azevedo and Leshno (2013). In large markets, agents act as price-takers but may still be able to manipulate outcomes by submitting a report that misrepresents their ordinal preferences (Azevedo and Budish, 2013).

Our empirical approach of considering strategic behavior is similar in spirit to He (2012) and Calsamiglia et al. (2014). He (2012) estimates preferences using data from the Boston mechanism in Beijing under the assumption that agents' reports are undominated. The set of undominated reports is derived using a limited number of restrictions implied by rationality, the specific number of schools and ranks that can be submitted in Beijing, and that the mechanism treats all agents symmetrically. The approach fully specifies the likelihood of reporting each of the undominated strategies. Calsamiglia et al. (2014) considers a model with strategic and non-strategic agents and fully specifies a likelihood for a boundedly rational decision maker. They model a strategic decision-maker that uses heuristics motivated by common strategic concerns in the Boston mechanism to pick the set of schools to rank before optimizing the list.

Compared to these previous approaches, we allow for a more general class of mechanisms that includes mechanisms with student priority groups. The proposed method does not require the researcher to analytically derive implications of rationality or pick ranking heuristics for estimation. Further, our aim is to characterize the identified set or show point identification under the restrictions imposed on the data and directly study the properties of an appropriate estimator, aspects which are not considered in these previous studies.

Previous research has questioned the extent to which agents are sophisticated. For example, Abdulkadiroglu et al. (2006) use particulars of the Boston mechanism to deduce reports that are clearly suboptimal and tabulate the fraction of agents that make one these reports. Recent evidence in Calsamiglia and Guell (2014) suggests that students in Barcelona responded to a change in strategic incentives when the system of assigning neighborhood priorities was administratively changed. We present a sharp condition for an agent's report to be consistent with equilibrium behavior that does not depend on details of the mechanism. This allows us to estimate the fraction of agents with reports that are not consistent with equilibrium behavior. It also shows that the equilibrium restriction we use in our approach is testable in the data. Extensions to formally test and/or relax the assumption of equilibrium behavior are left for future research.

We use techniques and build on insights from the identification of discrete choice demand (Matzkin, 1992, 1993; Lewbel, 2000; Berry and Haile, 2010). While the primitives are similar, unlike discrete choice demand, each report is a risky prospect that determines a probability of assignment to the schools on the list. Since choices over lotteries depend on expected utilities, our data contain direct information on cardinal utilities when the lotteries are not degenerate. In this sense, our paper is similar to Chiappori et al. (2012), although their paper focuses on risk attitudes rather than the value of underlying prizes.

This paper is related to the large, primarily theoretical, literature that has taken a mechanism design approach to the student assignment problem (Gale and Shapley, 1962; Shapley and Scarf, 1974; Abdulkadiroglu and Sonmez, 2003). Theoretical results from this literature has been used to guide redesigns of matching markets (Roth and Peranson, 1999; Abdulkadiroglu et al., 2006, 2009). While preferences are fundamental primitives that influence mechanism comparisons, prospective analysis of a proposed change in the school choice mechanism is rare (see Pathak and Shi, 2013, for an exception). A significant barrier is that the fundamental primitives are difficult to estimate since a large number of school choice mechanisms are susceptible to manipulation (Pathak and Sonmez, 2008, 2013). Results in this paper may allow such analysis. For instance, our techniques will allow comparing the welfare effects of a change to the Deferred Acceptance mechanism for a school district that uses the Boston mechanism. The relative benefits of these two mechanisms has been debated in the theoretical literature. Ergin and Sonmez (2006) show that full-information Nash equilibria of the Boston Mechanism are Pareto inferior to outcomes under the Deferred Acceptance mechanism. However, when analyzing Bayesian Nash Equilibria, stylized theoretical models with an assumed distribution of preferences have arrived at ambiguous conclusions about the welfare comparison between the two mechanisms (Miralles, 2009; Abdulkadiroglu et al., 2011; Featherstone and Niederle, 2011; Troyan, 2012). In the context of a multi-unit assignment problem, Budish and Cantillon (2012) use preferences solicited from a strategyproof mechanism for assigning courses to evaluate average assignment ranks under a manipulable mechanism.

Our methods may also be useful in extending recent work on measuring the effects of school assignment on student achievement that jointly specifies the preferences for schools and test-score gains (Hastings et al., 2009; Walters, 2013; Nielson, 2013). This work has been motivated by the fact that without data from a randomized assignment of students to schools, a researcher must account for sorting on unobservable preferences/characteristics that are also related to achievement gains. Additionally, estimates of preferences may be useful in extrapolating lottery based achievement designs if there is selection on the types of students that participate in the lottery (Walters, 2013). Methods for estimating preferences in a broader class of data-environments may expand our ability to study the effects of school assignment on student achievement.

This paper also contributes to the growing literature on methods for analyzing preferences in matching markets. Many recent advances have been made in using pairwise stability as an equilibrium assumption on the final matches to recover preference parameters (Choo and Siow, 2006; Fox, 2010b,a; Chiappori et al., 2011; Agarwal, 2013; Agarwal and Diamond, 2014). The data environment considered here is significantly different and pairwise stability need not be a good approximation for assignments from manipulable mechanisms. Another strand of the literature directly interprets agent behavior in matching markets in terms of preferences. For example, Hitsch et al. (2010) estimate preferences in an online dating marketplace that strategically avoid costs of emailing potential mates that are unlikely to respond. Similar considerations related to probability of success arise when applying to colleges and other search environments (Chade and Smith, 2006).

The proposed two-step estimator leverages insights from the industrial organization literature, specifically the estimation of empirical auctions (Guerre et al., 2000; Cassola et al., 2013) and dynamic games (Bajari et al., 2007). As in the methods used in those contexts, we use a two-step estimation procedure where the distribution of actions from other agents is used to construct probabilities of particular outcome as a function of the agents' own action and a second step is used to recover the primitives of interest. However, the nature of primitives, reports, the mechanism and economic environment are significantly different than in our context.

Overview

Section 2 describes the Cambridge Controlled Choice Plan and presents evidence that students are responding to strategic incentives provided by the mechanism. Section 3 sets up the model and notation for the results on identification and estimation. Section 4 presents the main insight of the paper on how to interpret submitted rank order lists. Section 5 presents the main convergence condition needed for our analysis, and describes and analyzes the class of report-specific priority + cutoff mechanisms. Section 6 studies identification under varying choice environments and the availability of a special regressor. Section 7 proposes a particular two-step estimator based on Gibbs' sampling. Section 8 applies our techniques to the dataset from Cambridge, MA. A reader interested in the empirical application instead of the econometric techniques may skip Sections 5-7. Section 9 concludes.

2 Evidence on Strategic Behavior

2.1 The Controlled Choice Plan in Cambridge, MA

We will use data from the Cambridge Public School's (CPS) Controlled Choice Plan for the academic years 2004-2005 to 2008-2009. Elementary schools in the CPS system assign about 41% of the seats through a partnerships with pre-schools (junior kindergarten or Montessori) or an appeals process for special needs students. The remaining seats are assigned through a "lottery process" that takes place in January for students entering kindergarten. We will focus on students and seats that are allocated through this process.

[Table 2 about here.]

Table 2 summarizes the students and schools. The CPS system has 13 schools and about 400 students participating in it each year. One of the schools, Amigos, was divided into bilingual Spanish and regular programs in 2005. Bilingual Spanish speaking students are considered only for the bilingual program, and students that are not bilingual are considered only for the regular program.[4] King Open OLA is a Portuguese immersion school/program that any student may apply to. Tobin, a Montessori school, divided admissions for four and five year olds starting 2007.

One of the explicit goals of the Controlled Choice Plan is to achieve socio-economic diversity by maintaining the proportion of students who qualify for the federal free/reduced lunch program in each school close to the district-wide average. Except Amigos and only for the purposes of the assignment mechanism, all schools are divided into paid lunch and free/reduced lunch programs. Students eligible for federal free or reduced lunch are only considered for the corresponding program.[5] About 34% of the students are on free/reduced lunch. Each program has a maximum number of seats and the overall school capacity may be lower than the sum of the seats in the two programs. Our dataset contains both the total number of seats in the school as well as the seats available in each of the programs.

The Cambridge Controlled Choice Mechanism

We now describe the process used to place students at schools. The process prioritizes students at a given school based on two criteria:

1. Students with siblings who are attending that school get the highest priority.

[4] A student voluntarily declares whether she is bilingual on the application form.

[5] Households with income below 130% (185%) of the Federal Poverty line are eligible for free (reduced) lunch programs. For a household size of 4, the annual income threshold was approximately \$29,000 (\$41,000).

2. Students receive priority at the two schools closest to their residence.

Students can submit a ranking of up to three programs at which they are eligible. A variant of the Boston Mechanism assigns students as follows:

Step 0: Draw a single lottery number for each student

Step $\boldsymbol{k} \in$ ***1,2,3:*** Each school considers all students that have not been previously assigned and have listed it in the k-th position and arranges them in order of priority, breaking ties using the lottery number. Starting from the first student in the list, students are considered sequentially:

- The student under consideration is assigned to the paid lunch program if she is not eligible for a federal lunch subsidy and there is an open seat in both the paid lunch program and the school. If she is eligible for a federal lunch subsidy, then she is assigned to the free/subsidized lunch program as long as seats are remaining in both the free lunch program and the school.

2.2 Descriptive Evidence on Ranking Behavior

Panels A and D in Table 3 show that over 80% of the students rank the maximum allowed number of schools and over 80% of the students are assigned to their top ranked choice in a typical year. Researchers in education have interpreted similar statistics in school districts as indicators of student satisfaction and heterogeneity in student preferences. For instance, Glenn (1991) argues that school choice caused improvements in the Boston school system based on observing an increase in the number of students that were assigned to their top choice.[6] Similarly, Glazerman and Meyer (1994) interpret a high fraction of students getting assigned to their top choice in Minneapolis as indicative of heterogeneous student preferences.

[Table 3 about here.]

Conclusions based on interpreting stated preferences as truthful are suspect when a mechanism provides strategic incentives for students. It is well understood that students risk "losing their priority" if a school is not ranked at the top of the list in mechanisms like the Boston mechanism (Ergin and Sonmez, 2006). Table 3, Panel E shows that students tend to rank schools where they have priority closer to the top. For instance, schools where a student has sibling priority is ranked first 32% of the time as compared to 35% of the time

[6] The argument is based on ranking and assignment data generated when Boston used a manipulable assignment system.

anywhere on the list. Likewise, schools where a student has proximity priority are also more likely to be ranked higher. These statistics do not necessarily indicate that this behavior is in response to strategic incentives because having priority may be correlated with preferences. However, given that strategic incentives may also result in similar patterns, it may be incorrect to estimate preferences by treating stated lists as true preferences. For example, Panels D and of Tables 2 and 3 respectively show that the top-ranked school is closer than the average school, and closer than other ranked schools. One may incorrectly conclude that students have strong preferences for going to school close to home if proximity priority is influencing this choice.

2.3 Strategic Incentives in Cambridge

Table 4 takes a closer look at the strategic incentives for students in Cambridge. Panel A shows the frequency with which students rank the various school options, the capacity at the various schools and the rank, and priority type of the first rejected student in a school. Panels B and C present identical statistics, but split by free/reduced lunch status of students. The table indicates significant heterogeneity in the competitiveness of the schools. Baldwin, Haggerty, Amigos, Morse, Tobin, Graham & Parks, and Cambridgeport are competitive schools with many more students ranking them than there is capacity. Panel A indicates that a typical student would be rejected in these schools if she does not rank it as her top choice. Indeed, Graham and Parks rejected all non-priority students even if they had ranked it first in three of the five years. The other schools typically admit all students that were not assigned to higher ranked schools. Additionally, the competitiveness of schools differs by paid lunch status. While Graham & Parks is very competitive for students that pay for lunch, it did not reject any free/reduced lunch students that applied to it in a typical year.

[Table 4 about here.]

There are two other features that are worth noting. First, there are few instances where students were rejected at their second or third choices. This suggests that, in Cambridge at least, strategic incentives may be particularly important when considering which school to rank first. Second, several paid lunch students rank competitive schools as their second or third choice. This may appear hard to rationalize as optimal behavior. However, it should be noted that despite the large number of students ranking competitive schools second, these choices are often not pivotal, as evidenced by the extremely large number of students that are assigned to their top choice. Another possibility is that students are counting on back-up schools, either at the third ranked school, a private or a charter school in case they remain unassigned. We will discuss these findings after we present our estimates.

2.4 Strategic Behavior: A Regression Discontinuity Approach

We now present evidence that students are responding strategically when choosing which school to rank first. Our empirical strategy is based on the assignment of proximity priority in Cambridge. A student receives priority at the two closest schools to her residence. We can therefore compare the ranking behavior of students that are on either side of the boundary where the proximity priority changes. If students are not behaving strategically and the distribution of preferences are continuous in distance to school, we would not expect the ranking behavior to change discretely at this boundary. On the other hand, the results in Table 4 indicate that a sophisticated student risks losing her proximity priority at competitive schools if she does not rank it first. We now test whether students are responding to this strategic incentive.

[Figure 1 about here.]

Figure 1 presents bin-scatter plots of the probability of ranking a school against the distance from a boundary. In panels (a) through (d), students have proximity priority at schools to the left of the dashed line. Panel (a) shows that the probability that a student ranks a school first increases discontinuously at the proximity boundary. Further, the response to distance to school is also high. The jump at the boundary may be attenuated because a student can rank only one of the two schools she has priority as her top choice. In contrast to panel (a), we do not see a discrete jump at the proximity boundary for the probability a school is ranked second or third. This should be expected because we saw earlier that one's priority is unlikely to be pivotal in the second or third choices. Panel (d) focuses on the probability of ranking the second and third closest schools at the top rank. The discontinuity is particularly clear, with students more likely to rank the closer school, i.e. the one where they have proximity priority. Together, panels (a) and (d) strongly suggest that students are responding to the proximity priority in their decision of which school to rank first. We consider two placebo tests of our design. First, we repeated the exercise in panel (d) for the two closest schools. Note that a student has proximity priority at both schools. Reassuringly, we do not see a discrete change in behavior. Second, we repeated the exercise in panel (a) considering only the set of schools where a student does not have proximity priority. Again, we do not see a remarkable change in ranking behavior at the placebo boundary.

While these results suggest an aggregate change in ranking behavior as a function of proximity priority, there are two important caveats that must be noted. First, the results do not show that all students are responding to strategic incentives in the mechanism, or that their reports are optimal. In what follows, we will assume that all agents are sophisticated in their choice. Second, it is possible that the response is primarily driven by residential choices with which parents picking a home so that the student receives priority at a more preferred school. A full model that considers the joint decision residential and school choices is left for future research.

These results contrast with Hastings et al. (2009), who find that the average quality of schools ranked did not respond to a change in the neighborhood boundaries in the year the change took place. Assuming that students prefer higher quality schools, their finding suggests that students did not strategically respond to the change in incentives. As suggested by Hastings et al. (2009), strategic behavior may not be widespread if the details of the mechanism and the change in neighborhood priorities are not well advertised. Charlotte-Mecklenberg had adopted the lottery system just a year prior to their study and the district did not make the precise mechanism clear. In contrast, Cambridge's Controlled Choice Plan is published on the school district's website and has been in place for several years. These institutional features may account for the observed differences in the student behavior.

3 Model

We consider school choice mechanisms in which students are indexed by $i \in \{1, \ldots, n\}$ and schools indexed by $j \in \{0, 1, \ldots, J\} = S$. School 0 denotes being unmatched. Each school has q_j^n seats, with $q_0 = \infty$. We now describe how students are assigned to these seats, their preferences over the assignments, and the equilibrium behavior.

3.1 Assignment Mechanisms

School choice mechanisms typically use submitted rank-order lists and defined student priority types to determine final assignments. As is the convention in the school choice literature, let $R_i \in \mathcal{R}_i$ be a rank-order list, where jR_ij' indicates that j is ranked above j'.[7] Student, if they so choose, may submit a rank-order list that does not reflect their true preferences over schools. Let student i's priority type be denoted $t_i \in T$. In Cambridge, t_i defines the free-lunch type, the set of schools the student has proximity priority at and whether or not the student has a sibling in the school

A mechanism is usually described as an outcome of an algorithm that takes these rank-order lists and priorities as inputs. To study properties of a mechanism and our methods, it will be convenient to define a mechanism as a function that depends on the number of students n instead of the outcome of an algorithm.

[7]The set $\mathcal{R}_i$ may depend on the student's priority type t_i and may be constrained. For example, students in Cambridge can rank up to three schools, and programs are distinguished by paid-lunch status of the student.

Definition 1. *A* ***mechanism*** Φ^n *is a function* $(\Phi_1, \ldots, \Phi_n)$ *where*

$$\Phi_i^n : \mathcal{R}^n \times T^n \to \Delta S$$

such that for all $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$*, and* $t = (t_1, \ldots, t_n) \in T^n$*,*

$$\frac{1}{n} \sum_{i=1}^{n} \Phi_{ij}^n(R, t) \leq q_j^n.$$

In this notation, the $i - j$ component of $\Phi^n(R)$, denoted $\Phi_{ij}^n(R, t)$ is the probability that student i is assigned to school j. Hence, the outcome for each student is in the J-simplex ΔS. In the Cambridge school system, there is a lottery that breaks ties between students. Such lotteries are a common source of uncertainty faced by students.

3.2 Utilities and Preferences

We assume that student i's utility from assignment into program j is given by $V(z_{ij}, \xi_j, \epsilon_i)$, where z_{ij} is a vector of observable characteristics that may vary by program or student or both, and ξ_j and ϵ_i are (vector-valued) unobserved characteristics. Let

$$v_i = (v_{i1}, \ldots, v_{iJ})$$

be the random vector of indirect utilities for student i with density $f_V(v_{i1}, \ldots, v_{iJ} | \xi, z_i)$, where $\xi = (\xi_1, \ldots, \xi_J)$ and $z_i = (z_{i1}, \ldots, z_{iJ})$. We normalize the utility of not being assigned, v_{i0}, to zero.[8]

This formulation allows for heterogeneous preferences conditional on observables. For instance, one may specify these indirect utilities as

$$v_{ij} = z_{ij}\beta_i + \xi_j + \varepsilon_{ij},$$

with parametric assumptions on the distribution of β_i, ξ_j, and/or $\varepsilon_i = (\varepsilon_{ij}, \ldots, \varepsilon_{iJ})$. The primary restriction thus far is that a student's indirect utility depends only on their own assignment and not of other students. This rules out preferences for peer groups or for convenience that carpool arrangements may afford.

[8]Scale normalizations needed for identification and estimation will be discussed in Section 6.

3.3 Equilibrium

We assume that agent behavior is described by a symmetric Bayesian Nash Equilibrium. Specifically, let $\sigma : \mathbb{R}^J \times T \to \Delta\mathcal{R}$ be a (symmetric) mixed strategy. The first argument of σ is the vector of utilities over the various schools, and the second argument is the priority type of the student. If a student forecasts that other students in the district are playing according to σ, her (ex ante) probability of assignment probability when she reports $R_i \in \mathcal{R}_i$ is given by the vector

$$\begin{aligned} L^{\sigma}_{n,R_i,t_i} &= \mathbb{E}_{\sigma}[\Phi^n((R_i,t_i),(R,T))|R_i,t_i] \\ &= \int \Phi^n((R_i,t_i),(R,T)) \prod_{k\neq i} \sigma_R(v_k,t_k)\mathrm{d}F_{V_{-i},T_{-i}}, \end{aligned} \tag{1}$$

where $\sigma_R(v,t)$ is the probability that an agent with utility vector v and priority type t reports R and $F_{V_{-i},T_{-i}} = \prod_{k\neq i} F_{V,T}$ is the distribution of utility and priority types of the other agents in the population. The (ex ante) probability of assignment therefore depends on both the lottery draw and the realization of the reports by the other students in the district.

Definition 2. *The strategy σ^* is a* ***type-symmetric Bayesian Nash Equilibrium*** *if $v \cdot L^{\sigma^*}_{n,R,t} \geq v \cdot L^{\sigma^*}_{n,R',t}$ for all $R' \in \mathcal{R}$ whenever $\sigma^*_R(v,t) > 0$.*

The focus on equilibrium play implies that students submit the report that maximizes their expected utility with correct notions of the distribution of play by other students. A student faces uncertainty due to both the distribution of reports that the other students will submit and due to uncertainty inherent in the mechanism. This approach contrasts with ex-post concepts of Nash Equilibria common in the literature on assignment mechanisms (see Ergin and Sonmez, 2006, for example). However, it is a natural starting point for analyzing mechanisms that are not dominant-strategy and is commonly taken in the empirical analysis of auction mechanisms (Guerre et al., 2000; Cassola et al., 2013, among others). Models of bounded rationality are beyond the scope of this paper.

Evidence presented in Section 2 suggests that agents are responding to strategic incentives in the Cambridge Mechanism. Further, anecdotal evidence suggests that parent groups and forums discussing ranking strategies are active (Pathak and Sonmez, 2008), and laboratory experiments suggests that strategic behavior is more common for manipulable mechanisms than strategyproof mechanisms (Chen and Sonmez, 2006; Calsamiglia et al., 2010). While direct evidence showing that agents play equilibrium strategies is limited, Calsamiglia and Guell (2014) observe a strategic response in the distribution of reports to a change in the allocation of neighborhood priorities. However, the assumption implies a strong degree of rationality and knowledge, particularly if parents vary in their level of sophistication as postulated by Pathak and Sonmez (2008, 2013).

4 A Revealed Preference Approach

This section illustrates the key insight of how we learn about the preferences of students from their (potentially manipulated) report, and present an overview of our method for estimating preferences.

Equation (1) reveals that a student's optimal choice depends on the (expected) assignment probabilities. These assignment probabilities depend on the strategies of the other agents, her report and her priority-type. In a Bayesian Nash Equilibrium, the expected assignment probabilities are consistent with the strategies of the other agents. The choice of a report by a student can be interpreted as a choice over the set of lotteries,

$$\mathcal{L}_{t_i}^{\sigma^*} = \left\{ L_{R_i,t_i}^{\sigma^*} : R_i \in \mathcal{R}_i \right\}.$$

These are the assignment probabilities that a student with priority type t_i can achieve by making different reports to the mechanism when the other agents are playing according to σ^*. We will suppress the dependence on σ^* and t_i in the notation for expositional simplicity, focusing on students with a given priority type and a Bayesian Nash Equilibrium.

Assume, for the moment, that the assignment probabilities available to a student is known to the analyst and consider her decision problem.[9] Figure 2 illustrates an example with two schools and an outside option. Each possible report corresponds to a probability of assignment into each of the schools and a probability of remaining unassigned. Figure 2(a) depicts three lotteries $L_R, L_{R'}, L_{R''}$ corresponding to the reports R, R' and R'' respectively on a unit simplex.[10] The dashed lines show the linear indifference curves over the lotteries for an agent with utility vector $v \in \mathbb{R}^J$. A student that is indifferent between L_R and $L_{R'}$ must have indifference curves that are parallel to the line segment connecting the two points and a utility vector $v_{R,R'}$ that is orthogonal to it. Figure 2(b) shows such utility vector. Likewise, students with a utility vector proportional to $v_{R,R''}$ are indifferent between L_R and $L_{R''}$. It is now easy to see that the shaded region in this figure denotes all utility vectors for which L_R is the optimal choice. More generally, for any J and set of lotteries $\mathcal{L}$, choosing

[9] The next section presents conditions under which the available data can be used to consistently estimate the assignment probabilities available to a student in the Bayesian Nash Equilibrium that generated the observed data.

[10] The simplex is often referred to as the Marschak-Machina triangle.

L_R is optimal if the utility vector belongs to the normal cone (or the polar dual):

$$C_R = \{v \in \mathbb{R}^J : v \cdot (L_R - L_{R'}) \geq 0 \text{ for all } R' \in \mathcal{R}\}.$$

For all values of v in this cone, the expected payoff from choosing R is at least as large as choosing any other report. Figure 2(c) illustrates the regions that correspond to R' and R'' being optimal choices in our example. It easy to see that the normal cones to any set of lotteries may intersect only at their boundaries, and together cover the utility space. Figure 2(d) shows this visually in our example. Specifically, in the space of utilities, the types $v_{R,R'}, v_{R,R''}$ and $v_{R',R''}$ are indifferent between two of the three choices. Reports R, R' and R'' are optimal for students with utility vectors in the regions C_R, $C_{R'}$ and $C_{R''}$ respectively.

[Figure 1 about here.]

The student's report therefore reveals which of the normal cones, $C_R \subseteq \mathbb{R}^J$ for $R \in \mathcal{R}$, contains her utility vector. We can use this insight to construct the likelihood of observing a given choice as a function of the distribution of utilities, f_V:

$$\mathbb{P}(R|z, \xi) = \int 1\{v \in C_R\} f_V(v|z, \xi) \mathrm{d}v. \tag{2}$$

This expression presents a link between the observed choices of the students in the market and the distribution of the underlying preferences, and will be the basis of our empirical approach. Note that the number of regions of the utility space that we can learn about from observed choices is equal to the number of reports that may be submitted to a mechanism, which grows rapidly with the number of schools or the number of ranks submitted.

There are three remaining issues to consider which we address in the subsequent sections. First, we propose a large class of mechanisms show for which the equilibrium assignment probabilities can be consistently estimated. This is essential for determining the regions C_R needed to construct the likelihood. The objective is to estimate the assignment probabilities for the equilibrium that generate the data, and therefore our procedure allows for multiple equilibria. Second, we provide conditions under which the distribution of utilities is non-parametrically identified. We can obtain point identification by "tracing out" the distribution of utilities with either variation in lottery sets faced by students or by using an additively separable student-school specific observable characteristic. Third, we propose a computationally tractable estimator based on Gibbs' sampling that can be used to estimate a parametric form for f_V. This estimator uses an estimate of the lotteries obtained from a first step.

5 A Class of Mechanisms and their Limit Properties

The first step of our procedure requires an estimate of the assignment probabilities. These probabilities are a result of mechanisms that are usually described in terms of algorithms that take in a profile of reports and priority types of all the students in the district. There are few a priori restrictions on these algorithms, allowing for mechanisms that may be ill-behaved. For instance, a small changes in number of students or their reports could potentially have large effects on the assignment probabilities.[11] Moreover, our objective is to estimate assignment probabilities simultaneously for all priority-types and each possible rank-order list that can be submitted by a student. These complications can create difficulties in obtaining precise estimates of assignment probabilities from the data.

This section presents a large class of mechanisms that have properties that allow for consistent estimation of assignment probabilities.

5.1 A Convergence Condition

To state our convergence condition, we first restrict attention to semi-anonymous mechanisms. These mechanisms treat students with the same priority type and report symmetrically. Formally,

Definition 3. Φ^n *is* ***semi-anonymous*** *if there exists a function* $\phi^n : (\mathcal{R}\times T)\times\Delta(\mathcal{R}\times T) \to \Delta S$*, such that*

$$\phi^n((R_i, t_i), m_{-i}) = \Phi^n_i((R_i, t_i), (R_{-i}, t_{-i})),$$

where $m_{-i} = \frac{1}{n-1}\sum_{i'\in -i}\delta_{(R_{i'},t_{i'})}$ *is the measure of reports of students other than* i.[12]

Semi-anonymous mechanisms use only the priority types and reports of students to determine assignments, and do not depend directly on the identities of the specific students. Therefore, only the number of reports made by each priority-type affect the final outcomes for each student. Additionally, a student's assignment probabilities only depends on the reported rank-order list and her priority type. The restriction that there are only finitely many priority types rules out a fine metric such as test scores that can be used to distinguish

[11] Two pathological examples allowed by Definition 1 are instructive. First, the assignment of all students depends only student 1's report. Second, an algorithm could depend on whether an odd or even number of students apply to schools.

[12] This definition is equivalent to the more usual definition: A mechanism is **semi-anonymous** with priorities T if (1) for all $R, t \in \mathcal{R}^n \times T^n$, and $i,\ i'$ such that $t_i = t_{i'}$, we have that $\Phi^n_i(R,t) = \Phi^n_{i'}(R,t)$ and (2) for all R_i, R_{-i} and permutations π of $-i = (1,\ldots,i-1,i+1,\ldots,n)$, we have that $\Phi^n_i((R_i,t_i),R_{-i},t_{-i}) = \Phi^n_i((R_i,t_i),R_{\pi(-i)},t_{\pi(-i)})$.

any two test scores.[13]

Our identification and estimation results are based on properties of the assignment probabilities in a large market. The key property that will allow us to proceed with the analysis for a mechanism is that outcomes of the mechanism evaluated at the empirical distribution of the reports converge in probability to the limiting values as the market grows in size. We state this condition formally as follows:

Condition 1 (Convergence at m). *Suppose the sequence of empirical measures m^{n-1} on $\mathcal{R} \times T$ converges in probability to the population measure $m \in \mathcal{M}$. Then, for each (R, t),*

$$|\phi^n((R,t), m^{n-1}) - \phi^\infty(R,t,m)| \xrightarrow{p} 0$$

where $\phi^\infty((R,t), m) = \lim_{n\to\infty} \phi^n((R,t), m)$.

This condition guarantees that if the distribution of reports and priority-types of other students converges to a limit m, then the sampling error in estimating the assignment probabilities using the observed sample vanishes as the sample size increases. It provides the basis for using the sample of reports observed for estimating assignment probabilities.

Specifically, consider assignment probabilities under samples with reports and priority types drawn from a sequence fo type-symmetric strategies $\sigma_R^n(v,t)$. These strategies may or may not be part of an equilibrium. We assume the sample of reports and priority types of the other players, m^{n-1}, is an empirical measure for a sample from

$$m^{\sigma^n}(R,t) = \int \sigma_R^n(v,t)\mathrm{d}F_{V,T}.$$

We now show that Condition 1 allows us to consistently estimate the assignment probabilities when the samples are generated from a sequence of type-symmetric strategies.

Theorem 1. *Assume that the sequence of type-symmetric strategies, σ^n, are such that $\|\sigma^n - \sigma\|_F \to 0$,[14] and ϕ^n satisfies Condition 1 at m^σ, then*

$$|\phi^n((R,t), m^{n-1}) - \phi^\infty((R,t), m^\sigma)| \xrightarrow{p} 0.$$

Proof. The proof follows from Condition 1. To apply this condition, we need to show that $\sup_{R,t} |m^{n-1}(R,t) - m^\sigma(R,t)| \xrightarrow{p} 0$. Note that $m^{n-1}(R,t)$ is a sample of $n-1$ independent

[13] Note that Φ_i^n only restricts $\phi^n((R_i,t_i), m(R_{-i}, t_{-i}))$ at a subset of probability measures m, namely, probability measures of the form $\frac{1}{n-1}\sum_i \delta_{R_i,t_i}$. We are free to choose ϕ^n at other values. Henceforth, we refer to a specific choice of ϕ^n when discussing a semi-anonymous mechanism.

[14] We use the norm $\|\sigma - \tilde{\sigma}\|_F = \sup_{R,t} \int |\sigma_R(v,t) - \tilde{\sigma}_R(v,t)|\mathrm{d}F_{V|t}$.

draws from $m^{\sigma^n}(R,t) = F_T(t)\int \sigma_R^n(v,t)\mathrm{d}F_{V|T}$. The triangle inequality implies that

$$\begin{aligned}
&\sup_{R,t} |m^{n-1}(R,t) - m^{\sigma}(R,t)| \\
\leq\ &\sup_{R,t} |m^{n-1}(R,t) - m^{\sigma^n}(R,t)| + \sup_{R,t} |m^{\sigma^n}(R,t) - m^{\sigma}(R,t)|.
\end{aligned}$$

The first term, converges in probability to 0 uniformly in R, t by the Glivenko-Cantelli theorem since $\mathcal{R} \times T$ is finite and therefore a universal Glivenko-Cantelli class. To show that the second term converges to zero, note that it can be rewritten and bounded using the triangle inequality as follows:

$$\begin{aligned}
&\sup_{R,t} \left| \int (\sigma_R^n(v,t) - \sigma_R(v,t))\mathrm{d}F_{V|t} \right| \\
\leq\ &\sup_{R,t} \int |\sigma_R^n(v,t) - \sigma_R(v,t)|\mathrm{d}F_{V|t} \\
=\ &\|\sigma_R^n - \sigma_R\|_F \to 0.
\end{aligned}$$

□

The condition above allows us to show that if σ^n is a convergent sequence of type-symmetric strategies, then the corresponding assignment probabilities converge to a limit. The condition therefore yields consistent estimates of assignment probabilities under alternative assumptions on behavior. Condition 1 is therefore agnostic about the solution concept and is best seen as a regularity condition guaranteeing consistent estimation of assignment probabilities. The techniques developed in this section may therefore be useful for extensions in which students need not be best responding to correct beliefs about assignment probabilities.

For our preference estimates, we will assume that student behavior is described by an equilibrium. The result above shows that we can construct consistent estimates of agent beliefs using the observed sample of reports. For instance, one can assume that the data are generated from any sequence of BNE that converges to a point where Condition 1 is satisfied. Requiring a convergent sequence of BNE ensures that the equilibrium behavior of agents is well-behaved under the data generating process. Conditions that guarantee the existence of such a sequence are presented in Menzel (2012). These conditions are presented in terms of smoothness conditions on the best-response function at the equilibrium of the limit game (the game defined by ϕ^{∞}). Unfortunately, these are not easily mapped to primitives. Alternatively, we can assume a behavioral model in which agent reports are made according to a limit equilibrium with a continuum of agents (Kalai, 2004; Azevedo and

Budish, 2013).[15] The advantage of this approach is that it avoids analyzing sequences of equilibria to derive consistency results. It may also be a reasonable behavioral assumption in itself. Appendix A shows that the difference in payoffs to agents under these two solution concepts are not significant when there are a large number of agents. Specifically, we show that the limit of a sequence of BNE is a limit equilibrium and that all limit equilibria are approximate BNE.

In both approaches, aggregate uncertainty about the distribution of the reports disappears in the limit, although it is present in any finite BNE. This feature is not unique to our setting and is implied in any large game (Kalai, 2004; Menzel, 2012). We return to this point when proposing an estimator for assignment probabilities. Second, we allow for the possibility of multiple equilibria. The objective is to estimate assignment probabilities for the equilibrium that generated the data. We can achieve this objective because these are only a function of the distribution of reported preferences and priority types, which are observed.

Verifying Condition 1 may not be straightforward because matching mechanisms are usually described using algorithms instead of functions that take a measure of reports as inputs. Continuity or uniform convergence properties that often allow for econometric consistency are therefore not directly available. A representation of the mechanism as a function may be necessary before proceeding. The next subsection describes a large class of mechanisms in which the condition is satisfied.

5.2 Report-Specific Priority and Cutoff Mechanisms

This section introduces a class of mechanisms called report-specific priorities + cutoff mechanisms. These mechanisms admit a particular representation of how reports and priorities map into assignment probabilities.[16]

We consider mechanisms in which each student is assigned an eligibility score for each school, and the student is assigned to her highest ranked choice for which her eligibility score exceeds the school's cutoff. In symbols, given cutoffs, $p_1, \ldots, p_j$, we consider mechanisms that a student with eligibility scores $e_i = (e_{i1}, \ldots, e_{iJ})$ that submitted report R_i is assigned to school j if

$$D_j^{(R_i,e_i)}(p) = 1\{e_{ij} \geq p_j,\ jR_i0\} \prod_{j' \neq j} 1\{jR_ij' \text{ or } e_{ij'} < p_{j'}\}. \tag{3}$$

We now describe how student eligibility scores and the school-specific cutoffs are determined.

[15]We say that σ^* is a **Limit Equilibrium** if $\sigma_R^*(v,t) > 0$ implies that $v \cdot \phi^\infty((R,t), m^{\sigma^*}) \geq v \cdot \phi^\infty((R',t), m^{\sigma^*})$ for all $R' \in \mathcal{R}$.

[16]Our representation is for any set of reports, not only for those generated from an equilibrium.

As the name suggests, eligibility scores in report-specific priority + cutoff mechanisms are modified based on the report made by the student and the priority type. Formally, we assume that there is a lottery ν_i that is not known to a student at the time the student makes her report. Let $\gamma_{\nu|t}$ denote the distribution of the lottery given the priority type. The vector eligibility scores is given by $e_i = f(R_i, \nu_i)$.

By allowing for the distribution of lotteries to depend on t, we allow for the case that sibling priority receive a more favorable distribution of lotteries than other students. We also allow for the distribution of the random priority to be correlated with the student's priority type and across schools. The dependence on f allows us to consider cases such as the Boston mechanism or First Preferences First prioritize all students that rank a school first over other students.

Finally, the allocations are determined by a school-specific cutoff $p_j \in [0,1]$. The cutoff, p_j, will be determined as a function of reports, priorities and lottery draws for all the students to ensure that schools are not assigned more students than there are positions available. Let $\eta \in \Delta\left(\mathcal{R} \times [0,1]^J\right)$ be a measure of student reports, and eligibility scores. We can now write the measure of students that are eligible for j and rank it above other eligible schools:

$$D_j(p|\eta) = \eta\left(\{e_{ij} \geq p_j,\ jR_i0\} \bigcap_{j' \neq j} \left(\{jR_ij'\} \cup \{e_{ij'} < p_{j'}\}\right)\right).$$

Given $D(p|\eta)$ and school capacities q, we can define the set of cutoffs that clear the market as follows:

Definition 4. *The vector of cutoffs p is a **market clearing cutoff** for economy (η, q) if for all $j \in S$, $D_j(p|\eta) - q_j \leq 0$, with equality if $p_j > 0$.*

At a market clearing cutoff, the total number of students that are eligible and seek assignment at any given school is no higher than the capacity at the school. Moreover, a school has a strictly positive cutoffs only if assigning students to their highest ranked choice for which they are eligible would exhaust the school's capacity. We consider mechanisms that use market clearing cutoffs to determine who is assigned to any given school.

Formally, we say that a mechanism ϕ^n is a **report-specific priorities + cutoff** mechanism if there exists a function $f : \mathcal{R} \times [0,1]^J \to [0,1]^J$ and a measure $\gamma_{\nu|t}$ over $[0,1]^J$ for each $t \in T$ such that

(i) $\phi^n_j((R_i, t_i), m(R_{-i}, t_{-i}))$ is given by

$$\int \cdots \int D^{(R_i, f(R_i, \nu_i))}(p^n) \mathrm{d}\gamma_{\nu_1|t_1} \ldots \mathrm{d}\gamma_{\nu_n|t_n}$$

where $f(R_i, \nu_i))$ is the modified priority,

(ii) p^n are market clearing cutoffs for capacity q^n and $\eta^n = \frac{1}{n} \sum_i \delta_{(R_i, f(R_i, \nu_i))}$.

(iii) f strictly increasing in the last J arguments.

The representation highlights two ways in which these mechanisms can be manipulable. First, the report of an agent can modify her eligibility. Fixing a cutoff, agents may have the direct incentive to make reports that may not be truthful. Second, even if eligibility does not depend on the report, an agent may (correctly) believe that the cutoff for a school will be high, making it unlikely that she will be eligible. If the rank-order list is constrained in length, she may choose to omit certain competitive schools.

This representation extends the characterization of stable matchings by Azevedo and Leshno (2013) in terms of demand-supply and market clearing to discuss mechanisms. Particularly, we can use the framework to consider mechanisms that produce matchings that are not stable. As we show in the next section, a remarkable feature of this representation is that it encompasses a very broad class of mechanisms that differ essentially by the choice of f. The representation may therefore be of independent theoretical interest.

5.2.1 Examples

This subsection shows that most commonly used mechanisms can be expressed as report-specific lottery + cutoff mechanisms. The main text focuses on the two most commonly used mechanisms:

The **Student Proposing Deferred Acceptance Mechanism**: For reports $R_1, \ldots, R_N$ and priorities $t_1, \ldots, t_N$,

Step* 1*: Students apply to their first listed choice and their applications are *tentatively* held in order of priority and a lottery number until the capacity has been reached. Schools reject the remaining students.

Step* k*: Students that are rejected in the previous round apply to their highest choice that has not rejected them. Schools pool new applications with those held from previous steps, and *tentatively hold* applications in order of priority and lottery number until capacity has been reached. The remaining students are rejected. The algorithm continues if any rejected student has not been considered at all their listed schools. Otherwise, each student is assigned to the school that currently holds her application.

This mechanism is strategy-proof for the students if the students can rank all J schools (Dubins and Freedman, 1981; Roth, 1982), but provides strategic incentives for students if students are constrained to list $K < J$ schools (see Abdulkadiroglu et al., 2009; Haeringer and Klijn, 2009, for details).

The **Boston Mechanism**: For reports $R_1, \ldots, R_N$ and priorities $t_1, \ldots, t_N$, each school

Step* 1*: Assign students to their first choice in order of priority and a lottery number until the capacity has been reached. Reject the remaining students.

Step* k*: Assign students that are rejected in the previous round to their k-th choice in order of priority and a lottery number until the capacity has been reached. Schools reject the remaining students. Continue if any rejected student has not been considered at all their listed schools.

This mechanism is a canonical example for one that provides strategic incentives to students (Abdulkadiroglu et al., 2006).

Proposition 1. *The Deferred Acceptance Mechanism and the Boston Mechanism with lotteries are report-specific priority + cutoffs mechanisms.*

Proof. See Appendix B.3. We use $e = f(R, \nu) = \nu$ for deferred acceptance and $e_j = f_j(R, \nu) = \dfrac{\nu_j - \#\{k : kRj\}}{J} + \dfrac{J-1}{J}$ for the Boston Mechanism. This choice of f for Boston upgrades the priority of the student at her first choice relative to all students that list that school lower. □

Remark 1. *Serial Dictatorship, First Preferences First, Chinese Parallel Mechanism and the Pan London Admissions scheme are also report-specific priority + cutoff mechanisms. For completeness, we discuss these mechanisms in Appendix B.3.*

Hence, all mechanisms in Table 1 except the TTC and Cambridge mechanisms are report-specific priority + cutoffs mechanisms. As we discuss below, our convergence result will require an additional assumption that the mechanism uses a lottery to break ties.

A researcher with data from one of these mechanisms will need to verify that priorities used by the mechanism satisfy our Assumptions above before applying the methods that follow. An important restriction is that the function f does not depend on the reports and priorities of the other agents. This may rule out some mechanisms that use the reports of other agents to determine eligibility in a program. Alternatively, one may prove Condition 1 directly, as we do for the Cambridge mechanism.

5.2.2 Condition 1 for Report-Specific Priority + Cutoff mechanism

Our main result in this section shows that this class of mechanism satisfy the key convergence condition needed to proceed with the rest of our analysis, and that this class contains the most commonly used mechanisms.

We make the following assumption on η in the limit continuum economy:

Assumption 1. *1. (Non-degenerate lotteries) For some $\kappa > 0$, and each p, $p' \in [0,1]^J$, $(R,t) \in \mathcal{R} \times T$ and j, $\eta_{e|R,t}(\{p_j \wedge p'_j \leq e_j \leq p_j \vee p'_j\}) \leq \kappa |p_j - p'_j|$.*

2. (Unique Cutoff) (η, q) admits a unique market clearing cutoff, p^*.

Non-degenerate lotteries is a strengthening of strict preferences in Azevedo and Leshno (2013). The assumption is straightforward to verify with knowledge of the mechanism. For example, it is satisfied if a lottery is used to break ties between multiple students with the same priority type. It also allows for a situation in which a single tie-breaking lottery that is used by all schools to break ties. This assumption, however, is not satisfied if the school district uses an exam with finitely many possible scores to determine eligibility and does not use a lottery to break ties between students with identical exam scores.

Assuming a unique cutoff restricts the joint distribution of reports and priorities, and the school capacities. Existence of a market clearing cutoff is guaranteed by Corollary A1 of Azevedo and Leshno (2013) for any η. Uniqueness is a restriction on an equilibrium object. Although the assumption is not made on primitives, it is a restriction on features that are observed in the data. Sufficient conditions that imply this assumption are therefore testable in principle. Further, using the reports observed in the data it is feasible to check if there are multiple cutoffs that approximately clear the market, but are sufficiently different. Not finding approximate market clearing cutoffs that are far might provide confidence in the assumption above. We refer the reader to Appendix B.1 for a more formal discussion of sufficient conditions for Assumption 1. This discussion borrows from results in Azevedo and Leshno (2013) and Berry et al. (2013).

We are now ready to state the first main result of this section.

Theorem 2. *Assume that (η, q) satisfies Assumptions 1, where*

$$\eta(\{R, e \leq p\}) = \sum_{t \in T} m(R,t)\gamma_{\nu|t}(\{f(R,\nu) \leq p\}).$$

ϕ^n satisfies Condition 1 if it is a report-specific priority + cutoff mechanism.

Proof. See Appendix B.2. □

The proof is based on a lemma showing that the market clearing cutoffs faced by an individual agent in the finite economy converges to the limiting cutoff p^*, irrespective of their lottery draw. This uniform convergence follows from standard empirical process results applied to the aggregate demand function. Intuitively, any single agent has a negligible effect on the aggregate demand function in the large market. We then use Assumption 1, which implies that the probability that a student with priority t_i and report R_i has an eligibility draw that is pivotal is negligible. Hence, the assignment probabilities in a large finite economy approach the limiting case.

An important feature of the representation of the mechanism in terms of the cutoffs and the use of these cutoffs in the proof is that is significantly reduces the dimensionality of the assignment probabilities that need to be estimated. The number of cutoffs is equal to the number of schools which is far fewer than $|\mathcal{R} \times T|$, the number of assignment probabilities that need to be estimated. This representation also implies that students only need to have correct beliefs about the (distribution of the) cutoffs in equilibrium. This is a lower dimensional object than assignment probabilities over which beliefs need to be formed.

6 Identification

In Section 4, we showed that the choice of report by a student allows us to determine the normal cone, $C_R \subseteq \mathbb{R}^J$ for $R \in \mathcal{R}$, that contains her utility vector v. This deduction required knowledge of the assignment probabilities L_R, which we showed can be consistently estimated under certain regularity conditions on the mechanism. We now articulate how one can learn about the distribution of utilities $f_{V|T}(v|z,\xi)$ by using implications of Equation 2:

$$\mathbb{P}(R \in \mathcal{R}|z,t,\xi,b) = \int 1\{v \in C_{b,R,t}\} f_{V|t}(v|z,\xi)\mathrm{d}v,$$

where b is a market subscript and the dependence on t has been reintroduced for notational clarity. It allows us to consider different market conditions for the same set of schools or students with different priority types.

The expression above shows that two potential sources of variation is available to the analyst that can be used to "trace out" the densities $f_{V|T}(v|z,\xi)$. First, we can consider choice environments with different values of $C_{b,R,t}$. Second we can consider variation in the observable characteristics z. We consider each of these in the subsequent sections. Appendix C.1 shows that the equilibrium assumption that reports are consistent with BNE is testable in principle under certain conditions on the mechanism.[17]

As is standard in the literature, our results consider the case when an infinitely large dataset is observed. Hence, the assignment probabilities and the fraction of students that

[17] The statistical testing problem is beyond the scope of this paper. Studying the problem may require proposing an alternative behavioral model and we suspect that it may involve testing for a parameter on the boundary.

choose any report are observed without sampling noise. We view these results as articulating the role of parametric assumptions as assisting estimation in finite samples.

6.1 Identification Under Varying Choice Environments

In some cases, a researcher is willing to exclude certain elements of the priority structure t from preferences, or may observe data from multiple years in which the set of schools are the same, but the capacity at schools varies across years. For instance, some students are grandfathered into Kindergarten from pre-K before the January lottery in Cambridge. This affects the number of seats available at a school during the lottery. This variation assists in identification if it is excluded from the distribution of utilities. This section illustrate what can be learned from such variation without any further assumptions.

[Figure 2 about here.]

When t is excluded from the distribution of preferences, i.e. $v|z, \xi, t \sim v|z, \xi, \tilde{t}$ for $t, \tilde{t} \in T$, we have students with the same distribution of preferences facing two different choice sets for assignment probabilities. For example, assume that the choice sets faced by t and $\tilde{t}$ are $\mathcal{L} = \{L_R, L_{R'}, L_{R''}\}$ and $\tilde{\mathcal{L}} = \{L_R, \tilde{L}_{R'}, L_{R''}\}$ respectively. Figure 3(a) illustrates these choice sets. The change from $L_{R'}$ to $\tilde{L_{R'}}$ affects the set of utilities for which the various choices are optimal. The set of types for with L_R is optimal is presented as the dotted cone. These utilities can be written as linear combinations of $\tilde{v}_{R,R'}$ and $v_{R,R'}$ with positive coefficients. Observing the difference in likelihood of reporting R for students with the two types allows us to determine the weight on this region:

$$\mathbb{P}(R|z, \tilde{t}) - \mathbb{P}(R|z, t) = \int (1\{v \in \tilde{C}_R\} - 1\{v \in C_R\}) f_{V|t}(v|z) dv.$$

Since utilities may be determined only up to positive affine transformations, normalizing the scale as $\|v_i\| = 1$ for each student i is without loss of generality. Hence, it is sufficiently to consider the case when $f_{V|t}$ has support only on the unit circle. Figure 3(b) illustrates that this variation allows us to determine the weight on the arc $\tilde{h}_R - h_R$. Appendix C.3 formally characterizes the identified set.

The discussion suggests that enough variation in the set of lotteries faced by individuals with the same distribution of utilities can be used to identify the preference distribution. If such variation is available, the arc above traces the density of utilities along the circle. Of course, we do not expect that typical variation in the data will be rich enough to use non-parametric estimation methods based on such variation. However, this observation articulates the sources of choice set variation that are implicitly used when utilities are not linked directly with priority types.

While this variation may not be rich enough for a basis for non-parametric identification, it makes minimal restrictions on the distribution of utility. In particular, the result allows for the distribution of utility to depend arbitrarily on residential locations. Although beyond the scope of this paper, this framework may be a useful building block for a model that incorporates both residential and schooling choices.

6.2 Identification With Preference Shifters

In this section we assume that the indirect utilities are given by

$$V(z_{ij}, \xi_j, \epsilon_i) = U(z_{ij}^2, \xi_j, \epsilon_i) - z_{ij}^1 \tag{4}$$

where $\epsilon_i \perp z_{ij}^1$. The magnitude of the coefficient on z^1 is most appropriately viewed as a scale normalization, and the model is observationally equivalent to one with random coefficient α_i that has support only on negative real numbers. This scale normalization replaces the normalization, $\|v_i\| = 1$, made in the previous section. For simplicity of notation, we will drop ξ, z^2 with the reminder that these are variables that the researcher needs to condition on. Let ζ be the support of z^1. Since $f_V(v|z^1)$ is a location family, this implies that $f_V(v|z^1) = g(v + z^1)$ where g is the density of u. Since the distribution of z^1 is observed in the data, our objective in this section is to identify the density g.

The term z_{ij}^1 is sometimes referred to as a special regressor (Lewbel, 2000; Berry and Haile, 2010, 2014). The linearly separable form and independence assumptions are the main restrictions in this formulation. In the school choice context, these assumptions needs to be made on a characteristic that varies by student and school. For instance, Abdulkadiroglu et al. (2014) assume that distance to school is independent of student preferences. The assumption is violated if unobserved determinants of student preferences simultaneously determine residential choices.

[Figure 3 about here.]

We now illustrate how variation in z^1 can be used to "trace-out" the density of u. Consider the lottery set faced by a set of students in Figure 2 and the corresponding region, C_R, of the utility space that rationalizes choice R. A student with $z^1 = z$ chooses R if $v = u - z \in C_R$. The values of u that rationalize this choice is given by $z + a_1 v_{R,R'} + a_2 v_{R,R''}$ for any two positive coefficients a_1 and a_2. Figure 4 illustrates the values of u that make R optimal. As discussed in Section 4, observing the choices of individuals allows us to determine the fraction of students with utilities in this set. Similarly, by focusing on the set of students with $z^1 \in \{z', z'', z'''\}$, we can determine the fraction of students with utilities in the corresponding regions. Figure 4 illustrates the sets that make R optimal for each of these values of z^1. By appropriately adding and subtracting the fractions, we can learn the fraction of students with utilities in the parallelogram defined by $z - z' - z''' - z''$. This allows us to learn the total weight placed by the distribution g on such parallelograms of arbitrarily small size. It turns out that we can learn the density of g around any point z in the interior of ζ by focusing on local variation around z. The next result formalizes this intuition.

Theorem 3. *Suppose C_R is spanned by J linearly independent vectors $\{w_1, \ldots, w_J\}$. If $h_{C_R}(z^1) = P(v \in C_R | z^1)$ is observed on an open set containing z^1, then $g(z^1)$ is identified. Hence, $f_V(v|z^1)$ is identified everywhere if $\zeta = \mathbb{R}^J$.*

Proof. Let $W = (w_1', \ldots, w_J')'$ be the matrix containing linearly independent vectors such that $C_R = \{v : v = Wa \text{ for some } a \geq 0\}$. Assume, wlog, $|\det W| = 1$. Evaluating h_{C_R} at Wx, we have that

$$h_{C_R}(Wx) = \int_{\mathbb{R}^J} 1\{u - Wx \in C\} g(u) \, \mathrm{d}u.$$

After the change of variables $u = Wa$:

$$\begin{aligned} h_{C_R}(Wx) &= \int_{\mathbb{R}^J} 1\{W(a-x) \in C_R\} g(Wa) da \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_J} g(Wa) \, da \end{aligned}$$

where the second inequality follows because $1\{W(a-x) \in C_R\} = 1\{a - x\} \leq 0$. Then:

$$\frac{\partial^J h_{C_R}(Wx)}{\partial x_1 \ldots \partial x_J} = g(Wx)$$

and $g(z^1)$ is given by $\frac{\partial^J h_C(Wx)}{\partial x_1 \ldots \partial x_J}$ evaluated at $x = W^{-1} z^1$. □

Intuitively, we use local changes in z^1 to shift the distribution of cardinal utilities to favor certain lotteries over others. Since simplicial cones are spanned by linearly independent vectors, we can decompose the change in how often a lottery is chosen into the principal directions to identify the density.

Note that the local nature of this identification result articulates precisely the fact that identification of the density at a point does not rely on observing extreme values of z^1. Of course, identification of the tails of the distribution of u will rely on support on extreme values of z^1. Also note that our identification result requires only one convex cone generated by a lottery, and therefore, observing additional lotteries with simplicial cones generates testable restrictions on the special regressor.

It turns out that considering cones C_R that are spanned by linearly independent vectors is sufficient for $J = 2$, but may not be useful for some sets of assignment probabilities if $J > 3$. This is because for $J = 2$, the normal cone C_R is spanned by linearly independent vectors if L_R is extremal (in the convex hull of $\mathcal{L}$). Intuitively, an extremal lottery can have only two other adjacent lotteries and therefore the cone C_R is spanned by two vectors. However, when $J > 2$, a lottery may have more than J adjacent lotteries, resulting in a cone C_R that is spanned by more than J vectors. These vectors cannot be linearly independent.

Fortunately, we can still identify g if z^1 has full support on $\mathbb{R}^J$ as long as the tails of g are exponentially decreasing. Theorem C.3 in Appendix C.4 states the results and conditions formally. The proof is based on Fourier-deconvolution techniques since the distribution of v is given by a location family parametrized by z^1. This allows us to learn about g from observing how choices over lotteries change with z^1. However, because the result is based on deconvolution techniques, it requires stronger support assumptions than in Theorem 3. Nonetheless, the conditions on $\mathcal{G}$ are quite weak, and are satisfied for commonly used distributions with additive errors such as normal distributions, generalized extreme value distributions or if u has bounded support.[18]

7 Estimation

Non-parametric estimation of random utility models can be computationally prohibitive and imprecise in finite samples, particularly if there are several options. Following the discrete choice literature, we parametrize the distribution of indirect utilities $F_V(v|z,\xi)$ with $F_{V;\theta}(v|z,\xi)$ where θ belongs to a compact set $\Theta \in \mathbb{R}^K$. This parametric representation is a parsimonious approximation to the identified primitives that is specifically chosen to answer a particular economic question. The identification results in the previous section show that these parametric assumptions may be relaxed in the presence of richer data.

We consider a two-step estimator where in the first step we replace $\phi^\infty((R,t),m)$ with a consistent estimate $\hat{\phi}(R,t)$. For example, $\hat{\phi}(R,t) = \phi^n((R,t),m^{n-1})$ where m^{n-1} is the empirical measure on the reports and priority types of $n-1$ agents in the sample. Condition

[18]We do not require that g has a non-vanishing characteristic function. When u has bounded support, the support conditions on ζ can also be relaxed. In this case, we can allow for ζ to be a corresponding bounded set.

1 implies that $\hat{\phi}(R,t) \xrightarrow{p} \phi^{\infty}((R,t),m)$. Our second step is defined as an extremum estimator:

$$\hat{\theta} = \inf_{\theta \in \Theta} Q_n(\theta, \hat{\phi})$$

Consistency of such a procedure is straightforward to establish under mild conditions on Q_n. The result is formally stated and proves in Appendix D.1.

The objective function Q_n could be based on a likelihood or a method of moments. We will implement our second-step as a Gibbs' sampler, and interpret the posterior mean of this sampler as asymptotically equivalent to the Maximum Likelihood Estimator. We now describe each of the steps for the Cambridge Mechanism and the particular parametric specification used in the second step.

7.1 First Step: Estimating Assignment Probabilities

The first step requires a (consistent) estimate of the assignment probabilities $\phi((R,t),m)$ as function of the reports and priority types, (R,t). Given Condition 1, there are several feasible methods for obtaining consistent estimates. For instance, one may use the observed assignment probabilities conditional on the ranks and priority types of the students. A significant disadvantage of this method is that several feasible rank-order lists may not be observed for a given priority-type, or may not be observed frequently enough to obtain accurate estimates.

Our preferred method is to simulate the mechanism directly and resample other students for each rank and priority type from the observed data. This uses the knowledge of the details of the mechanism and avoids the small sample size problem that a method that uses the observed assignments confronts. While one may simply use the observed reports of the other students, we believe that resampling the other students is likely to better approximate the uncertainty the students face in finite samples. In our dataset, we implement this by categorizing students into various types and iterating through feasible rank order lists. For each list, we use 1,000 draws of the lotteries and $N-1$ other students (drawn with replacement) along with their observed rank-order lists and priority types.

A final possibility is to take advantage of the representation of mechanisms as report-specific priority + cutoff mechanism and directly simulate the cutoffs. Then, for each rank-order list, one may compute the probability of assignment for each student. This can alleviate computational difficulties in simulating the mechanism when the number of feasible rank order lists or the number of priority types is large. The complexity grows exponentially in number of schooling options because one iterates through various rank lists. In contrast, the cutoffs grow linearly in the number of schools, this method may ease computation.[19] We cannot apply this method because the Cambridge mechanism is not a Rank Specific Priority + Cutoffs mechanism, although it satisfies Condition 1.

7.2 Second Step: Preference Estimates

While our identification results do not make parametric assumptions on utilities, we implement the following parametric specification to assist estimation in finite samples. Student i's utility school j is:

$$\begin{aligned} v_{ij} &= \sum_{k=1}^{K} \delta_{kj} x_{ik} - d_{ij} + \varepsilon_{ij} \\ v_{i0} &= 0 \end{aligned} \tag{5}$$

where d_{ij} is the road distance between student i's home and school j; x_{ik} are indicators of student demographic characteristics such as ethnicity, home language, and sibling priority; δ_{kj} are school specific parameters to be estimated; $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ}) \sim N(0, \Sigma)$.[20] The normalization of $u_{i0} = 0$ is without loss of generality, and the scale normalization is embedded in the assumption that the coefficient on d_{ij} is -1.

We specify independent prior distributions for $\delta = \{\delta_{jk}\}_{j=1..J,k=1..K} \in \mathcal{R}^{JK}$ and Σ. It is convenient to use a normal prior on δ, $\delta \sim N(\overline{\delta}, A^{-1})$ and an independent inverse Wishart prior on Σ,$\Sigma \sim IW(\nu_0, V_0)$. We follow McCulloch and Rossi (1994) and use a Gibbs' sampler to obtain draws of δ and Σ from the posterior distribution exploiting the fact that the chosen priors are conditionally conjugate.

We construct a chain of draws $(v^s, \delta^s, \Sigma^s)_{s=1}^{S}$ using a Gibbs Sampler. The initial conditions v^0, Σ^0 are set such that for every student i, $v_i^0 \in C_{R_i}$, where C_{R_i} is the cone associated with R_i, the observed report chosen by i; and Σ^0 is a random draw from $IW(\nu_0, V_0)$.

For each draw s, the sampler uses v^{s-1} to obtain the parameters δ of equation (5). We use the standard procedure in Bayesian approaches to draw δ^s from the posterior distribution of δ given its prior, the data (v^{s-1}, x) and the distribution of error terms $N(0, \Sigma^{s-1})$. A new draw Σ^s is drawn from the posterior distribution of Σ given its prior and the data ϵ^s which consists on the error terms implied by equation (5) when $v = v^{s-1}$ and $\delta = \delta^s$. The last step in draws v^s. We set $v^s = v^{s-1}$ and update each of its elements iterating over students and

[19] Lemma B.1 in the appendix shows consistency of the cutoffs estimated from the data for a Rank Specific Priority + Cutoff mechanism.

[20] Note that our specification allows for heteroskedastic errors ε_{ij} and arbitrary correlation between ε_{ij} and $\varepsilon_{ij'}$. This specification relaxes homoskedastic and independent preference shocks commonly used in logit specifications.

schools. For student i and school j, we draw v_{ij}^s from a truncated normal distribution with mean determined by δ^s, variance determined by Σ^s and truncation points determined by $v_{i,-j}^s$.[21] This procedure ensures that $v_i^s \in C_{R_i}$ for every student i in every step s. Additional details on the implementation of our Gibbs' sampler are in Appendix D.2.

The sampler can be initialized at any starting value for Σ and v as long as $v_i \in C_{R_i}$ for all i. Values of v_i consistent with the constraint $v_i \in C_{R_i}$ can be obtained using well known linear programming techniques. Notice that an empty solution set to $v_i \in C_{R_i}$ is equivalent to a choice that cannot be rationalized as being strictly optimal.

8 Application to Cambridge

8.1 Estimated Assignment Probabilities

Table 5 presents the lottery estimates. As in Table 4, the estimates indicate considerable heterogeneity in school competitiveness. The typical student isn't guaranteed assignment at the more competitive schools even if she ranks it first. On the other hand, several schools are sure shots for students that rank them first. The probability of not getting assigned to a school also differs with paid lunch status. A comparison of estimates in Panel A with those in Panels D and E indicates that having priority at a school significantly improves the chances of assignment. The differential is larger if the school is ranked first.

[Table 5 about here.]

Perhaps one surprising feature is that the estimated probability of assignment is zero in very few cases. Indeed, paid lunch students ranking Graham & Parks as the second choice, or one of Graham & Parks, Haggerty or Baldwin as the third choice are the only cases in which the probability of assignment is estimated to be zero. Table 4 might have suggested that it students are much less likely to be assigned to the latter two schools if they rank it second. One reason for this difference is that the calculation in Table 5 accounts for uncertainty in the set of students that are drawn. Although this uncertainty vanishes in the large market, the calculation that resamples students from the observed data may better approximate the uncertainty perceived by students if they do not know exactly what the other students report.

[21] Our problem is therefore slightly is different from, although not more difficult than, a Gibbs' sampler approach to estimating standard discrete choice models in McCulloch and Rossi (1994). The standard discrete choice models only involve sampling from one-sided truncated normal distributions.

8.2 Preference Estimates: Truthful vs Sophisticated Players

We compute the posterior distribution of preference parameters using the set of students that submitted a rank-order list consistent with optimal play (i.e. submitted a list corresponding to an extremal lottery). A total of 1,958 students (92% of the sample) submitted a rationalizable list.[22] The large fraction of students with rationalizable lists may initially appear surprising. However, Theorem C.1 indicates that the lists that are not rationalized are likely the ones where assignment probabilities for one of the choices is zero. Our estimates in Table 5 suggest that this is rare, except for a few schools. Indeed, most of the students with lists that cannot be rationalized listed Graham & Parks as their second choice. Therefore, one concern with dropping students with lists that cannot be rationalized is that we are liable to underestimate the desirability of competitive schools.

[Table 6 about here.]

Panel A of Table 6 presents the (normalized) mean utility for various schools net of distance, by student group for two specifications. The first specification treats the agent reports as truthful, while the second treats all agents as sophisticated. The underlying parameter estimates for the model with sophisticated agents are presented in Appendix Table F.2. In both specifications, we find significant heterogeneity in willingness to travel for the various school options. Paid lunch students, for instance, place a higher value on the competitive schools as compared to the non-competitive schools. Although not presented in the mean utilities, Spanish and Portuguese speaking students disproportionately value schools with bilingual and immersion programs in their home language. Students also place a large premium on going to school with their siblings.

The estimates suggest that treating stated preferences as truthful may lead to underestimates of the value of competitive schools relative to non-competitive schools. This differential is best illustrated using Graham & Parks as an example. Treating stated preferences as truthful, we estimate that paid lunch students have an estimated mean utility that is an equivalent of 1.26 miles higher than the average public school option. This is an underestimate relative to the model that treats agents as sophisticated. In contrast, the value of Graham & Parks for free lunch students is over-estimated by the truthful model. Specifically, treating agents as sophisticated reveals that it is less desirable than the typical public school option for the average student. The difference can be explained by observing that Graham & Parks is not competitive for free lunch students, and therefore, the low number of applications it receives indicates particular dislike for the school from this group of students.

[22] One student was dropped because the recorded home address data could not be matched with a valid Cambridge street address.

Another significant difference between the two sets of estimates is the number of schools students find preferable to the outside option. Panel B shows that estimates that treat stated preferences as truthful suggest that more than half the students have five or more schools where assignment is preferable to the outside option. On the other hand, treating agents as sophisticated suggests that more than half the students find assignment at three or fewer schools in the system preferable to the outside option. Treating preferences as truthful extrapolates from the few students (about 13%) that do not have complete rank order lists. On the other hand, the model that treats students as sophisticated interprets the decision to rank long-shots in the second and third choices as evidence of dislike for the remaining schools relative to the outside option. These results should be viewed in light of Cambridge's thick after-market. About 92.56% of the students that are not assigned though the lottery choice process are assigned to one of the schools in the system. In fact, more than a quarter of unassigned students are placed at their top ranked school through the wait-list. There are also charter school and private school options that unassigned students may enroll in. The value of the outside option is therefore best interpreted in terms of the inclusive value of participating in this after-market.[23]

8.3 Ranking Behavior and Assignment

In this section we investigate the ranking strategy of agents, whether they would suffer large losses from out-of-equilibrium truth-telling, and how strategic manipulation may affect student welfare.

Table 7 presents the fraction of students that find truthful reporting optimal and losses from truthful behavior as estimated using the two assumptions on student behavior. The first three columns can be interpreted as analyzing the true loss to students from not behaving strategically if they are indeed out-of-equilibrium truth-tellers. The estimates suggest that the truthful report is optimal for 56% of the students. The average student suffers a loss equal to 0.19 miles by making a truthful report, or 0.43 miles conditional on regretting truthful behavior. The losses are heterogeneous both within and across student groups. Free-lunch students, for instance, suffer losses from truthful play less often and suffer lower losses conditional on any losses. This reflects the fact that the Cambridge school system is not competitive for these students because of the seats specifically reserved for them.

[Table 7 about here.]

[23]To some extent, students that are assigned through the process can choose to enroll elsewhere, should there be open seats. This may question the interpretation of the mean utility estimates for the inside options. However, approximately 91% of the students that are assigned through the lottery process enroll in their assigned school.

The last three columns estimate preferences assuming that agents are sophisticated but tabulate losses from non-strategic behavior. Again, these estimates suggest that a little less than half the students, and disproportionately paid-lunch students have strategic incentives to manipulate their reports. Together, the observations suggest that markets where students face large competitive pressures are precisely the markets where treating preferences as truthful may lead to biased assessments of how desirable various schools are.

[Table 8 about here.]

Our estimates that about half the students find it optimal to behave truthfully is likely to affect our assessment of how many students are assigned to their top choice. Table 8 presents this fraction by student paid lunch status. The last column indicates that, on average, 88% of the students rank their top choice first. This occurs because many students avoid ranking competitive schools as their top rank in favor of increasing the odds of assignment to a less preferred option. As a result, fewer students rank Graham & Parks as their top choice, instead favoring Haggerty or Baldwin. We therefore see over-subscription to Haggerty and Baldwin by paid lunch students relative to the true first choice. Consequently, the last column indicates that while 84% were assigned to their stated first choice, only 75% were assigned to their true first choice. This pattern is particularly stark for paid lunch students, who are assigned to their true first choice only 68% of the time. Table 5 indicated that assignment to competitive schools is less likely for paid lunch students. Together, these results suggest that calculations of whether students are assigned to their preferred options based on stated preferences may be misleading, and differentially so by student demographics.

8.4 Welfare Effects of Redesigning the Mechanism

A central question in the mechanism design literature is whether variants of the Boston Mechanism are worse for student welfare as compared to strategy-proof mechanisms such as the Deferred Acceptance Mechanism. This question has been debated in the theoretical literature with stylized assumptions on the preference distribution (see Miralles, 2009; Abdulkadiroglu et al., 2011; Featherstone and Niederle, 2011). The Boston mechanism exposes students to the risk that they are not assigned to their top listed choices, which can harm welfare when they strategically choose not to report their most preferred schools. However, this risk has a countervailing force that only agents with particularly high valuations for their top choice will find it worthwhile listing the schools on top. Hence, the mechanism screens agents for cardinal preferences and can increase aggregate welfare.

Table 9 presents two different analyses of this question. The first assumes that students report their preferences truthfully to the Cambridge mechanism, while the second implements our proposed method that treats agents as sophisticated. In both cases, we compare the outcomes to the Student Proposing Deferred Acceptance Mechanism.[24] An approach that treats agents' stated preferences in the Cambridge Mechanism as truthful finds little difference in the average welfare between the two mechanisms. This estimate may be biased if strategic behavior is widespread.

[Table 9 about here.]

In contrast, the results that treat agents as sophisticated indicate that the Cambridge mechanism outperforms the Deferred Acceptance mechanism. Particularly, both the fraction of students assigned to their true first choice choice and the average utility of students is higher under the Cambridge mechanism. Paid lunch students would benefit more from such a change than free/reduced lunch students. However, only half the students prefer the Cambridge Mechanism to the Deferred Acceptance Mechanism. This observation suggests that the mechanism is effectively screening based on cardinal utilities. The magnitude of the difference between the Cambridge Mechanism and the Deferred Acceptance Mechanism is similar to the difference between Deferred Acceptance and Pareto Efficient Matching in New York City, as measured by Abdulkadiroglu et al. (2014). We do not advocate for the Cambridge mechanism based on our results for two reasons. First, only about half the students prefer the Cambridge Mechanism to Deferred Acceptance. Second, there may be welfare consequences to strategic behavior due to strategizing or distributional consequences to naive behavior.

Our quantitative results contribute to the debate in the theoretical literature about the welfare properties of the Boston mechanism, which is similar to the Cambridge mechanism. The results are different in spirit from Ergin and Sonmez (2006), that suggests that full-information Nash equilibria of the Boston Mechanism are Pareto inferior to outcomes under Deferred Acceptance. This difference stems from our focus on Bayesian Nash Equilibria that accounts for uncertainty faced by the students. Abdulkadiroglu et al. (2011) theoretically show that the Boston Mechanism can effectively screens for the intensity of preferences and can have better welfare properties than the Deferred Acceptance Mechanism. Troyan (2012) shows that the theoretical results in this literature that are based on notions of interim efficiency are not robust to students having priorities, and advocates for an ex-ante comparison such as the one performed in this paper.

[24] We construct the school choice function by adapting the Cambridge Controlled Choice Plan. Schools consider students according to their total priority + lottery. A paid lunch student's application is held if the total number of applications in the paid lunch category is less than the number of available seats and if the total number of held applications is less than the total number of seats. Free lunch student applications are held in a similar manner. We allow students to rank all available choices.

Evaluating the counterfactual market in which the student proposing deferred acceptance mechanism is adopted is relatively straightforward. This is because the mechanism is strategy-proof. Evaluating a counterfactual mechanism that is manipulable may be harder. However, it may be easier to solve for equilibrium cutoffs for a rank-specific priority + cutoff mechanism and use the cutoffs to compute assignment probabilities. This approach allows for analysis of counterfactual mechanisms that are manipulable.

It is important to note that there may be distributional consequences if agents vary in their ability to strategiz (Pathak and Sonmez, 2008). Additionally, agents may face costs when strategizing since it may require students to gather additional information before formulating ranking strategies. Our calculations ignore these costs, which weigh against using Boston-like mechanisms for school assignment.

9 Conclusion

We develop a general method for analyzing preferences from reports made to a single unit assignment mechanism that may not be truthfully implementable. We view the choice of report as a choice from available assignment probabilities. The available probabilities can be consistently estimated under a weak condition on the convergence of a sequence of mechanism to a limit. The condition is verified for a broad class of lottery-based school choice mechanisms including the Boston mechanism or the Deferred Acceptance mechanism. We then characterize the identified set of preference distributions under the assumption that agents play a Bayesian Nash Equilibrium. The set of preference distributions are typically not point identified, but may be with sufficient variation in the lottery set. We then obtain point identification if a special regressor is available.

The methods in this paper rely on sophisticated agents participating in the mechanism. Ranking behavior in Cambridge indicates that agents respond to the strategic incentives in the mechanism. Specifically, students that reside on either side of the boundary where proximity priority changes have observably different ranking behavior. We take this as evidence against the assumption that agents are ranking schools according to true preferences. We then implement our method using the proposed estimator. Our estimates indicate that treating preferences as truthful is likely to result in biased estimates in markets where students face stiff competition for their preferred schools. The stated preferences therefore exaggerate the fraction of students assigned to their top choice. We also illustrate how our method can be used to evaluate changes in the design of the market. Specifically, we find that the typical student would be worse-off by an equivalent of 0.1 miles if Cambridge switched to the student proposing deferred acceptance mechanism. These losses are concentrated at the paid lunch students, who are assigned to their top choice less often under the Deferred Acceptance mechanism. This calculation, however, ignores potential distributional consequences of heterogeneous agent sophistication and costs of strategizing.

Our methods can be extended in several directions. In the context studied here, schools are passive players who express their preferences with only coarse priorities and a random lottery. Extending the techniques to allow for exam scores may be of interest to allow for finely defined priority groups may allow for broader applicability of the results, but may require technical innovations for estimating the assignment probabilities. Another important extension is to consider a college admissions setting where students make optimal applications while considering their probabilities of admissions. A challenge in directly extending our approach is that we observe all priorities relevant for admissions in the data. In the college applications context, admissions may depend on unobservables that also affect preferences, complicating the analysis. A closely related context is a multi-unit assignment mechanism such as course allocation mechanisms where agents play truthfully. The preferences in this context would need to be richer in order to allow for complementarities over the objects in a bundle that are assigned to an individual. Last, but not least, relaxing the assumptions on the sophistication of agents is an important avenue for future research with several possible alternative approaches to consider. These extensions are interesting avenues for expanding our ability to analyze agent behavior in assignment mechanisms.

References

Abdulkadiroglu, A., Agarwal, N., and Pathak, P. A. (2014). The Welfare Effects of Coordinated School Assignment: Evidence from the NYC High School Match.

Abdulkadiroglu, A., Che, Y.-K., and Yasuda, Y. (2011). Resolving Conflicting Preferences in School Choice: The Boston Mechanism Reconsidered. *American Economic Review*, 101(1):399–410.

Abdulkadiroglu, A., Pathak, P. A., and Roth, A. E. (2009). Strategy-Proofness Versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match.

Abdulkadiroglu, A., Pathak, P. A., Roth, A. E., and Sonmez, T. (2006). Changing the Boston School Choice Mechanism. *NBER Working Papers 11963*.

Abdulkadiroglu, A. and Sonmez, T. (2003). School Choice: A Mechanism Design Approach. *The American Economic Review*, 93(3):729 – 747.

Agarwal, N. (2013). An Empirical Model of the Medical Match.

Agarwal, N. and Diamond, W. F. (2014). Identification and Estimation in Two-sided Matching Markets.

Ayaji, K. (2013). School Choice and Educational Mobility: Lessons from Secondary School Applications in Ghana.

Azevedo, E. and Budish, E. (2013). Strategyproofness in the large.

Azevedo, E. and Leshno, J. (2013). A Supply and Demand Framework for Two-Sided Matching Markets.

Bajari, P., Benkard, C. L., and Levin, J. (2007). Estimating Dynamic Models of Imperfect Competition. *Econometrica*, 75(5):1331–1370.

Beggs, S., Cardell, S., and Hausman, J. (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, 17(1):1–19.

Berry, S. T., Gandhi, A., and Haile, P. A. (2013). Connected Substitutes and Invertibility of Demand. *Econometrica*, 81(5):2087–2111.

Berry, S. T. and Haile, P. A. (2010). Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers.

Berry, S. T. and Haile, P. A. (2014). Identification in Differentiated Products Markets Using Market Level Data. *Econometrica, forthcoming.*

Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841 – 890.

Berry, S. T., Levinsohn, J., and Pakes, A. (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy*, 112(1):68–105.

Block, H. and Marshak, J. (1960). Random Orderings and Stochastic Theories of Responses. In Olkin, I., Ghurye, S., Hoeffding, W., Mado, W., and Mann, H., editors, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press.

Budish, E. and Cantillon, E. (2012). The Multi-unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard. *American Economic Review*, 102(5):2237–2271.

Calsamiglia, C., Fu, C., and Guell, M. (2014). Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives.

Calsamiglia, C. and Guell, M. (2014). The Illusion of School Choice: Empirical Evidence from Barcelona.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained School Choice: An Experimental Study. *American Economic Review*, 100(4):1860–74.

Cassola, N., Hortaçsu, A., and Kastl, J. (2013). The 2007 Subprime Market Crisis Through the Lens of European Central Bank Auctions for Short Term Funds. *Econometrica*, 81(4):1309–1345.

Chade, H. and Smith, L. (2006). Simultaneous Search. *Econometrica*, 74(5):1293–1307.

Chen, Y. and Kesten, O. (2013). From Boston to Chinese parallel to deferred acceptance: Theory and experiments on a family of school choice mechanisms. *Discussion Papers, Research Unit: Market Behavior*.

Chen, Y. and Sonmez, T. (2006). School choice: an experimental study. *Journal of Economic Theory*, 127(1):202–231.

Chiappori, P.-A., Salanie, B., Salanie, F., and Gandhi, A. (2012). From Aggregate Betting Data to Individual Risk Preferences.

Chiappori, P.-A., Salanié, B., and Weiss, Y. (2011). Partner Choice and the Marital College Premium.

Choo, E. and Siow, A. (2006). Who Marries Whom and Why. *Journal of Political Economy*, 114(1):175–201.

Dubins, L. and Freedman, D. (1981). Machiavelli and the Gale-Shapley Algorithm. *The American Mathematical Monthly1*, 88(7):485–494.

Ergin, H. and Sonmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics2*, 90(1-2):215–237.

Featherstone, C. and Niederle, M. (2011). School choice mechanisms under Incomplete Information: an Experimental Investigation.

Fox, J. T. (2010a). Estimating Matching Games with Transfers.

Fox, J. T. (2010b). Identification in matching games. *Quantitative Economics*, 1(2):203–254.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Glazerman, S. and Meyer, R. H. (1994). Public school choice in Minneapolis, midwest approaches to school reform. *Proceedings of a Conference Held at the Federal Reserve Bank of Chicago*.

Glenn, C. L. (1991). Controlled choice in Massachusetts public schools. *Public Interest*, 103:88–105.

Guerre, E., Perrigne, I., and Vuong, Q. (2000). Optimal Nonparametric Estimation of First-Price Auctions. *Econometrica*, 68(3):525–574.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947.

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous Preferences and the Efficacy of Public School Choice.

He, Y. (2012). Gaming the Boston School Choice Mechanism in Beijing. mimeo, Columbia University.

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2010). Matching and Sorting in Online Dating. *American Economic Review*, 100(1):130–163.

Kalai, E. (2004). Large Robust Games. *Econometrica*, 72(6):1631–1665.

Kojima, F. and Pathak, P. A. (2009). Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review*, 99(3):608 – 27.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1):145–177.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, 8(3):229–254.

Matzkin, R. L. (1992). Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models. *Econometrica*, 60(2):239–70.

Matzkin, R. L. (1993). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics*, 58(1-2):137–168.

McCulloch, R. and Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. pages 105 – 142.

Menzel, K. (2012). Inference for large games with exchangeable players.

Miralles, A. (2009). School Choice: The Case for the Boston Mechanism.

Nielson, C. (2013). Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students.

Pathak, P. A. and Shi, P. (2013). Simulating Alternative School Choice Options in Boston.

Pathak, P. A. and Sonmez, T. (2008). Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism. *American Economic Review*, 98:1636–1652.

Pathak, P. A. and Sonmez, T. (2013). School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation. *American Economic Review*, 103(1):80–106.

Pennell, H., West, A., and Hind, A. (2006). Secondary School Admissions in London.

Roth, A. E. (1982). The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7(4):617–628.

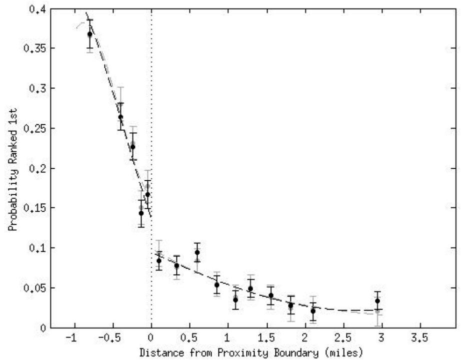
Roth, A. E. and Peranson, E. (1999). The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, 89(4):748–780.

Shapley, L. and Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, 1(1):23–37.

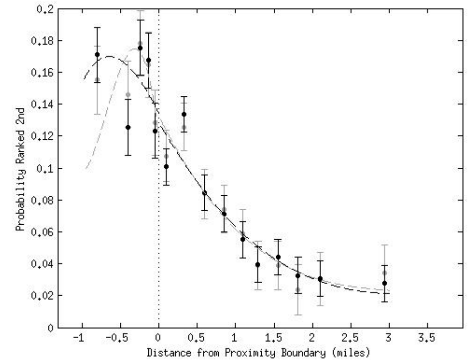
Troyan, P. (2012). Comparing school choice mechanisms by interim and ex-ante welfare. *Games and Economic Behavior*, 75(2):936–947.

Walters, C. (2013). A Structural Model of Charter School Choice and Academic Achievement.

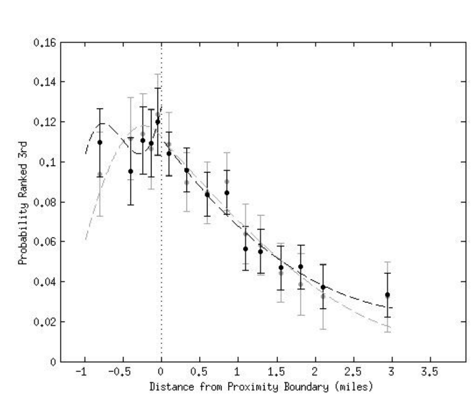
Figure 1: Effect of Proximity Priority on Ranking Behavior



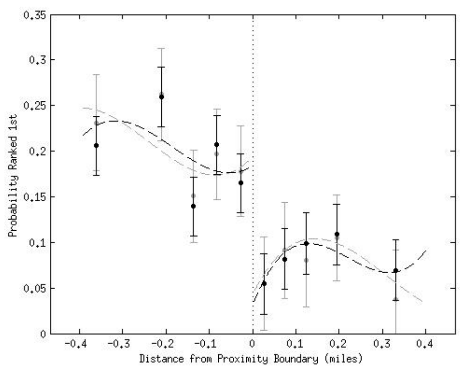
(a) First Rank



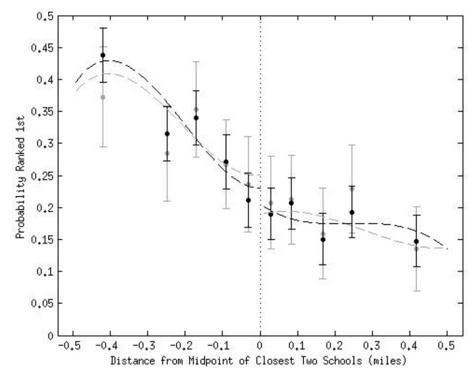
(b) Second Rank



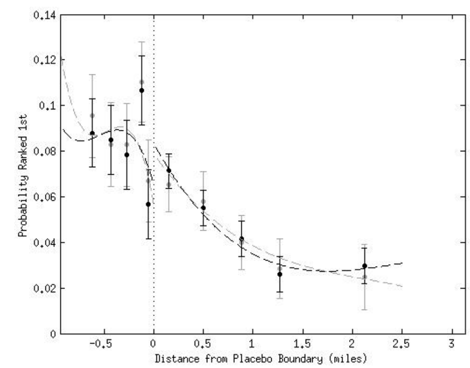
(c) Third Rank



(d) Second and third closest schools



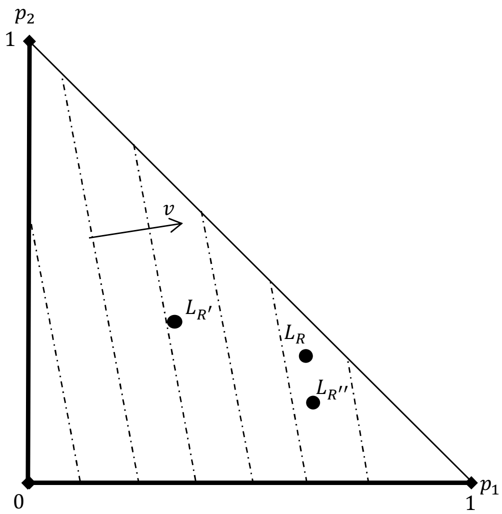
(e) Placebo at prioritized schools



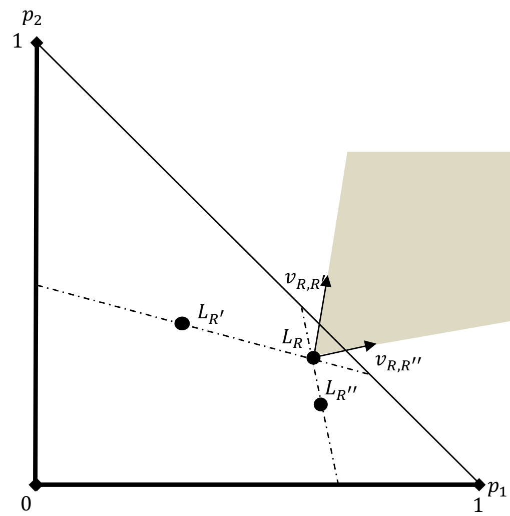
(f) Placebo at unprioritized schools

Notes: The graphs are bin-scatter plots (based on distance) with equally sized bins on either side of the boundary. For each student, we construct a boundary distance, $\bar{d}_i$, based on her distance to the schooling options. For a given school-student pair, the horizontal axis represents $d_{ij} - \bar{d}_i$. The vertical axis is the probability that a student ranks the school in the relevant distance bin. Range plots are 95% confidence intervals. Black plot points are raw data, which the grey points control for school fixed effects. Dashed lines represent cubic fits. Panels (a) through (d) use the average distance between the second and third closest schools as the boundary. A student is given proximity priority at the schools to the left of the boundary and does not receive priority at schools to the right. Panel (d) considers only the second and third closest schools. Panel (e) uses the average distance between the closest and second closest schools. Only schools where students have proximity priority are considered. Panel (f) uses the average distance between the fourth and fifth closest schools, and only the schools that a student does not have proximity priority. Panels (a), (d), (e) and (f) plot the probability that a school is ranked first. Panels (b) and (c) plot the probability that a school is ranked second and third respectively. Distances as calculated using ArcGIS. Proximity priority recorded by Cambridge differs from these calculations in about 20% of the cases. Graphs are qualitatively similar when using only students with consistent calculated and recorded priorities. Details in data appendix.

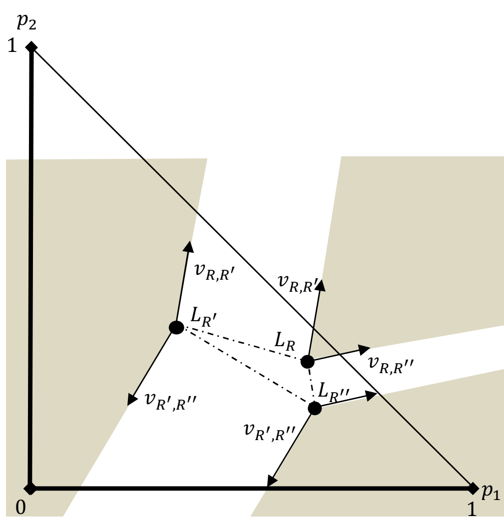
Figure 2: A Revealed Preference Argument



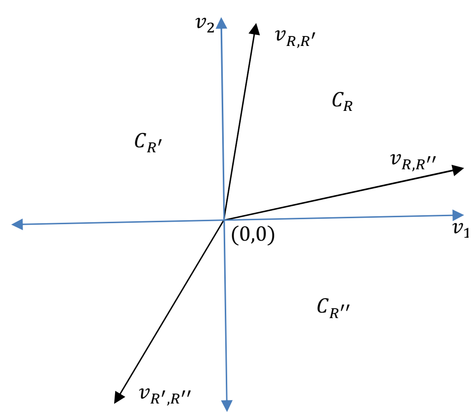
(a) Indifference curves for utility vector v, and choices over three lotteries



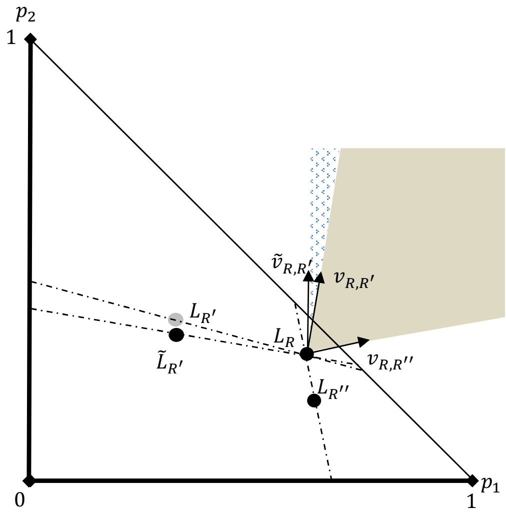
(b) L_R is optimal for v in the normal cone (shaded region), which is given by $v = a_1 v_{R,R'} + a_2 v_{R,R''}$ for $a_1, a_2 > 0$



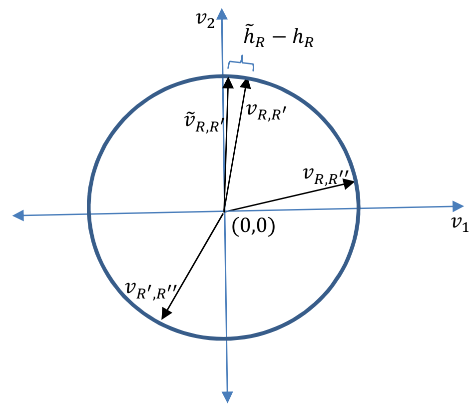
(c) Normal cones partition utility space



(d) Lottery choice reveals utility region



(a) Variation in the assignment probabilities induces variation in the normal cones



(b) Allows identifying the density of utility distribution on the unit circe

Figure 3: Variation in Lotteries

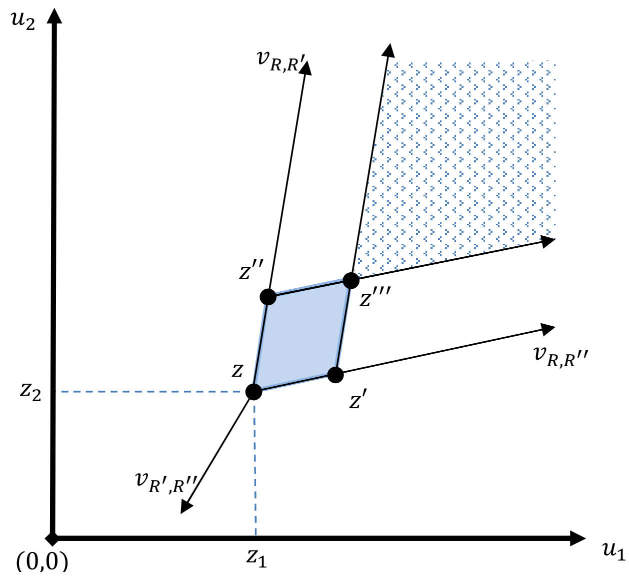


Figure 4: Local variation in z identifies the density of u

Table 1: School Choice Mechanisms

Mechanism	Manipulable	Examples
Boston Mechanism	Y	Barcelona[1], Beijing[2], Boston (pre 2005), Charlotte-Mecklenberg[3], Chicago (pre 2009), Denver, Miami-Dade, Minneapolis, Seattle (pre 1999 and post 2009), Tampa-St. Petersburg.
Deferred Acceptance		
w/ Truncated Lists	Y	New York City[4], Ghanian Schools, various districts in England (since mid '00s)
w/ Unrestricted Lists	N	Boston (post 2005), Seattle (1999-2008)
Serial Dictatorships		
w/ Truncated Lists	Y	Chicago (2009 onwards)
First Preferences First	Y	various districts in England (before mid '00s)
Chinese Parallel	Y	Shanghai and several other Chinese provinces[5]
Cambridge	Y	Cambridge[6]
Pan London Admissions	Y	London[7]
Top Trading Cycles		
w/ Truncated Lists	Y	New Orleans[8]

Notes: Source Table 1, Pathak and Sonmez (2008) unless otherwise stated. See several references therein for details. Other sources: [1] Calsamiglia and Guell (2014); [2]He (2012); [3]Hastings et al. (2009); [4]Abdulkadiroglu et al. (2009); [5]Chen and Kesten (2013); [6] "Controlled Choice Plan" CPS, December 18, 2001; [7]Pennell et al. (2006);
[8]http://www.nola.com/education/index.ssf/2012/05/new_orleans_schools_say_new_pu.html accessed May 20, 2014.

Table 2: Cambridge Elementary Schools and Students

Year	2004	2005	2006	2007	2008	Average
			Panel A: District Characteristics			
Schools	13	13	13	13	13	13
Programs	24	25	25	27	27	25.6
Seats	473	456	476	508	438	470
Students	412	432	397	457	431	426
Free/Reduced Lunch	32%	38%	37%	29%	32%	34%
Paid Lunch	68%	62%	63%	71%	68%	66%
			Panel B: Student's Ethnicity			
White	47%	47%	45%	49%	49%	47%
Black	27%	22%	24%	22%	23%	24%
Asian	17%	18%	15%	13%	18%	16%
Hispanic	9%	11%	10%	9%	9%	10%
			Panel C: Language spoken at home			
English	72%	73%	73%	78%	81%	76%
Spanish	3%	4%	4%	4%	3%	3%
Portuguese	0%	1%	1%	1%	1%	1%
			Panel D: Distances(miles)			
Closest School	0.43	0.67	0.43	0.47	0.45	0.49
Average School	1.91	1.93	1.93	1.93	1.89	1.92

Notes: Students participating in the January Kindergarten Lottery. Free/Reduced lunch based on student's application for Federal lunch subsidy.

Table 3: Cambridge Elementary Schools and Students

Year	2004	2005	2006	2007	2008	Average
	Panel A: Round of assignment					
First	81%	84%	85%	83%	75%	82%
Second	8%	3%	4%	7%	5%	5%
Third	5%	2%	2%	2%	4%	3%
Unassigned	6%	11%	9%	8%	16%	10%
	Panel B: Round of assignment: Paid Lunch Students					
First	80%	77%	78%	79%	68%	76%
Second	5%	4%	5%	8%	5%	5%
Third	6%	3%	4%	2%	3%	4%
Unassigned	9%	16%	14%	11%	24%	15%
	Panel C: Round of assignment: Free Lunch Students					
First	85%	95%	98%	94%	89%	92%
Second	14%	1%	2%	4%	6%	5%
Third	2%	1%	0%	1%	4%	1%
Unassigned	0%	4%	0%	2%	1%	1%
	Panel D: Number of Programs Ranked					
One	2%	6%	9%	5%	12%	7%
Two	5%	6%	9%	7%	7%	7%
Three	93%	89%	82%	88%	81%	87%
	Panel E: Ranking Schools with Priority					
Sibling at Ranked School	42%	38%	33%	26%	36%	35%
Sibling at 1st Choice	38%	34%	32%	24%	34%	32%
Sibling Priority at 2nd Choice	4%	3%	1%	2%	2%	2%
Sibling Priority at 3rd Choice	0%	2%	1%	1%	0%	1%
Proximity at 1st Choice	53%	52%	50%	51%	52%	51%
Proximity at 2nd Choice	42%	34%	37%	33%	37%	36%
Proximity at 3rd Choice	22%	24%	24%	25%	21%	23%
	Panel F: Distance (miles)					
Ranked first	1.19	1.18	1.24	1.29	1.19	1.22
All ranked schools	1.37	1.41	1.38	1.40	1.34	1.38
Assigned School	1.10	1.01	1.07	1.12	0.92	1.04

Notes: Proximity priority as reported in the Cambridge Public School assignment files.

Table 4: School Popularity and Competitiveness

School	Graham Parks	Haggerty	Baldwin	Morse	Amigos	Cambridgeport	King Open	Peabody	Tobin	Flet Mayn	Kenn Long	MLK	King Open Ola
						Panel A: All Students							
Ranked First	60	56	53	47	37	34	33	31	25	18	16	12	5
Ranked Second	72	37	66	25	18	44	39	38	17	10	18	20	0
Ranked Third	56	33	46	31	19	44	37	32	20	15	16	15	0
Ranked Anywhere	192	120	166	102	75	113	114	105	64	48	54	51	6
Capacity	41	41	41	42	41	27	51	48	35	38	41	37	15
First Rejected	1-P	1-R	1-R	1-R	1-R	1-R	NR	NR	1-R	NR	NR	NR	NR
						Panel B: Paid Lunch Students							
Ranked First	49	45	40	29	25	24	25	17	13	4	7	4	2
Ranked Second	60	28	56	14	12	29	23	27	10	3	6	6	0
Ranked Third	47	29	33	19	15	34	24	18	11	4	8	10	0
Ranked Anywhere	152	95	128	60	51	87	70	65	33	9	21	20	3
Capacity	29	27	27	29	41	18	36	34	29	35	34	27	15
First Rejected	1-P	1-R	1-R	1-R	1-R	1-R	NR	NR	3-R	NR	NR	NR	NR
						Panel C: Free Lunch Students							
Ranked First	9	12	12	17	12	11	13	10	12	16	10	9	2
Ranked Second	13	8	7	11	5	12	17	12	8	8	14	11	0
Ranked Third	10	4	9	10	4	12	13	13	9	10	11	4	0
Ranked Anywhere	29	24	25	40	20	36	44	38	31	36	34	25	2
Capacity	25	23	26	26	41	17	33	31	19	18	26	24	15
First Rejected	NR	NR	NR	1-R	1-R	2-P	NR	NR	1-R	NR	NR	NR	NR

Notes: Median number of applicants and seats over the years 2004-2008. First rejected is the round and priority of the first rejected student, e.g., 1-P indicates that a student with proximity priority was rejected in the first round. S: Sibling priority, PS: both proximity and sibling priority, R: regular/no prioirity, and NR: no student was rejected in any round. Free/Reduced lunch based on student's application for Federal lunch subsidy.

Table 5: Estimated Assignment Probabilities

Ranked	Graham Parks	Haggerty	Baldwin	Morse	Amigos	Amigos Bilingual	Cambridgeport	King Open	Peabody	Tobin	Tobin K4	Tobin K5	Flet Mayn	Kenn Long	MLK	King Open Ola
Panel A: All Students																
First	0.44	0.58	0.62	0.57	0.73	0.98	0.51	1	0.95	0.86	0.28	0.35	0.92	1	1	1
Second	0.24	0.23	0.23	0.21	0.33	0.95	0.15	0.9	0.83	0.75	0.03	0.13	0.85	1	0.99	1
Third	0.16	0.15	0.13	0.09	0.2	0.78	0.08	0.59	0.54	0.62	0.02	0.06	0.72	0.84	0.83	0.85
Panel B: Paid Lunch																
First	0.23	0.44	0.48	0.53	0.73	1	0.42	1	0.94	0.93	0.28	0.37	1	1	1	1
Second	0	0.03	0.04	0.16	0.32	1	0.05	0.86	0.82	0.78	0.01	0.15	1	1	1	1
Third	0	0	0	0.07	0.2	0.83	0.01	0.52	0.54	0.64	0.01	0.08	0.86	0.86	0.86	0.86
Panel C: Free/Reduced Lunch																
First	0.85	0.86	0.89	0.64	0.73	0.97	0.7	1	0.95	0.72	0.28	0.28	0.77	1	1	1
Second	0.7	0.63	0.59	0.3	0.33	0.91	0.35	0.97	0.85	0.71	0.05	0.1	0.55	0.99	0.98	1
Third	0.49	0.44	0.4	0.15	0.2	0.73	0.21	0.73	0.56	0.59	0.03	0.03	0.44	0.8	0.77	0.81
Panel D: Proximity Priority																
First	0.65	0.97	0.95	0.92	0.94	0.99	0.89	1	0.99	0.92	0.55	0.55	0.95	1	1	1
Second	0.11	0.19	0.15	0.23	0.38	0.97	0.26	0.97	0.84	0.79	0.04	0.14	0.76	1	1	1
Third	0.08	0.08	0.05	0.11	0.24	0.83	0.11	0.69	0.61	0.64	0.02	0.09	0.63	0.87	0.84	0.88
Panel E: No Priority																
First	0.38	0.55	0.59	0.53	0.71	0.97	0.47	1	0.94	0.86	0.25	0.33	0.92	1	1	1
Second	0.28	0.23	0.23	0.21	0.32	0.96	0.15	0.89	0.83	0.76	0.02	0.13	0.87	1	0.99	1
Third	0.19	0.15	0.14	0.09	0.2	0.81	0.08	0.58	0.53	0.63	0.02	0.06	0.73	0.84	0.83	0.85

Note: Average estimates weighted by number of students of each type. Probabilities estimated from 1,000 simulations of the Cambridge mechanism. Ranks and priority types of opposing students are drawn with replacement from the observed data. Second and third rank assignment probabilities are conditional on no assignment to higher ranked choices, averaged across feasible rank order lists.

Table 6: Estimated Mean Utilities

	Truthful			Sophisticated		
	All Students	Paid Lunch	Free Lunch	All Students	Paid Lunch	Free Lunch
	Panel A: Mean Utility					
Graham Parks	0.96	1.26	0.35	0.9	1.44	-0.18
	[0.05]	[0.06]	[0.08]	[0.11]	[0.12]	[0.19]
Haggerty	1.13	1.36	0.67	1.02	1.3	0.45
	[0.06]	[0.07]	[0.11]	[0.11]	[0.11]	[0.19]
Baldwin	0.98	1.24	0.46	0.97	1.14	0.63
	[0.06]	[0.06]	[0.09]	[0.08]	[0.09]	[0.11]
Morse	0.64	0.63	0.66	0.82	0.79	0.89
	[0.06]	[0.07]	[0.08]	[0.09]	[0.10]	[0.11]
Amigos	-0.19	-0.06	-0.45	0.02	0.11	-0.16
	[0.12]	[0.13]	[0.15]	[0.21]	[0.21]	[0.24]
Cambridgeport	0.54	0.74	0.14	0.5	0.58	0.33
	[0.05]	[0.06]	[0.07]	[0.17]	[0.17]	[0.19]
King Open	0.54	0.63	0.36	0.51	0.6	0.32
	[0.06]	[0.06]	[0.07]	[0.08]	[0.09]	[0.10]
Peabody	0.28	0.19	0.45	0.03	0.04	0.02
	[0.07]	[0.08]	[0.09]	[0.10]	[0.10]	[0.13]
Tobin	-0.14	-0.52	0.6	-0.33	-0.69	0.37
	[0.10]	[0.11]	[0.12]	[0.18]	[0.20]	[0.21]
Flet Mayn	-0.87	-1.28	-0.05	-1.4	-2.01	-0.18
	[0.12]	[0.15]	[0.10]	[0.30]	[0.37]	[0.19]
Kenn Long	0	-0.21	0.43	-0.09	-0.24	0.19
	[0.07]	[0.09]	[0.07]	[0.14]	[0.15]	[0.15]
MLK	-0.44	-0.69	0.04	-0.7	-0.95	-0.19
	[0.09]	[0.10]	[0.09]	[0.13]	[0.15]	[0.13]
King Open Ola	-3.41	-3.28	-3.66	-2.24	-2.12	-2.49
	[0.32]	[0.31]	[0.41]	[0.33]	[0.35]	[0.38]
Outside Option	-1.96	-2.18	-1.52	-0.69	-0.54	-0.98
	[0.08]	[0.09]	[0.10]	[0.05]	[0.06]	[0.06]
	Panel B: Number of Acceptable Schools					
Up to 1	7%	5%	10%	6%	7%	3%
Up to 2	13%	10%	20%	24%	30%	14%
Up to 3	21%	16%	30%	53%	63%	36%
Up to 4	29%	23%	40%	75%	84%	58%
Up to 5	40%	34%	51%	87%	93%	75%

Notes: Panel A presents the average estimated utility for each school, normalizing the mean utility of the inside options to zero. Utilities calculated averaging over all students the predicted utility given their non-distance covariates and the estimated coefficients. Standard errors (standard deviation of the posterior distribution) in brackets. Panel B presents the cumulative distribution of the number of acceptable schools, i.e. schools that are preferred to the outside option, as implied by the posterior distribution of utilities.

Table 7: Losses from Truthful Reports

	Truthful						Sophisticated					
	No Loss		Mean Loss		Std Loss		No Loss		Mean Loss		Std Loss	
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
All	56%	0.01	0.19	0.02	0.55	0.05	55%	0.01	0.05	0.01	0.20	0.02
Free Lunch	67%	0.02	0.02	0.00	0.12	0.03	71%	0.02	0.01	0.00	0.08	0.02
Paid Lunch	51%	0.01	0.27	0.03	0.65	0.06	47%	0.02	0.07	0.01	0.23	0.03
Black	65%	0.02	0.07	0.02	0.32	0.06	68%	0.02	0.03	0.01	0.15	0.04
Asian	56%	0.03	0.21	0.04	0.58	0.08	55%	0.03	0.05	0.01	0.18	0.05
Hispanic	60%	0.03	0.11	0.03	0.38	0.10	57%	0.04	0.03	0.01	0.14	0.04
White	52%	0.01	0.26	0.03	0.64	0.06	49%	0.02	0.06	0.01	0.23	0.03
Other Race	46%	0.06	0.22	0.07	0.52	0.15	50%	0.06	0.07	0.03	0.19	0.07

Notes: Estimated loss from reporting preferences truthfully, relative to optimal report in distance units (miles).

Table 8: Ranking and Assignment of Top Choice

	Graham Parks	Haggerty	Baldwin	Morse	Amigos	Cambridgeport	King Open	Peabody	Tobin	Flet Mayn	Kenn Long	MLK	King Open Ola	Total
						Panel A: All Students								
Preferred School	20.6	11.2	9.2	10.9	7.9	6.7	7.7	7.3	5.9	4.0	4.0	2.5	1.1	98.9
Ranked #1 (simul)	15.4	12.1	10.7	11.4	8.3	7.5	9.2	8.3	5.5	4.0	4.0	2.6	1.1	100.0
Ranked #1 (data)	14.3	12.6	11.9	11.0	8.8	7.7	8.2	7.8	5.7	4.4	3.8	2.7	1.2	100.0
Preferred and Ranked #1	13.9	10.1	8.3	10.3	7.5	6.1	7.7	7.3	4.9	3.9	4.0	2.5	1.1	87.5
Preferred and Assigned	9.4	8.3	6.6	8.5	6.8	5.1	7.7	7.2	4.1	3.5	4.0	2.5	1.1	74.7
Ranked #1 and Assigned	10.0	9.7	8.3	9.4	7.4	6.1	9.2	8.1	4.3	3.6	4.0	2.6	1.1	83.7
						Panel B: Free Lunch Students								
Preferred School	8.7	7.1	6.5	12.9	7.1	6.6	7.4	7.9	10.5	10.7	7.5	4.8	1.6	99.4
Ranked #1 (simul)	8.9	7.6	6.9	13.0	7.3	6.9	7.6	8.3	9.0	10.5	7.6	4.9	1.6	100.0
Ranked #1 (data)	6.7	8.3	8.0	12.2	7.8	6.4	7.7	9.0	8.5	10.8	7.1	5.5	2.0	100.0
Preferred and Ranked #1	8.5	7.0	6.4	12.5	6.9	6.4	7.4	7.9	8.7	10.2	7.5	4.8	1.6	95.9
Preferred and Assigned	7.9	6.6	6.1	10.6	6.2	5.5	7.4	7.8	6.8	9.3	7.5	4.7	1.6	88.0
Ranked #1 and Assigned	8.2	7.1	6.5	11.0	6.5	5.9	7.6	8.2	6.9	9.5	7.6	4.9	1.6	91.4
						Panel C: Paid Lunch Students								
Preferred School	26.6	13.3	10.5	9.9	8.2	6.7	7.9	7.0	3.5	0.7	2.2	1.4	0.8	98.6
Ranked #1 (simul)	18.8	14.4	12.7	10.6	8.8	7.8	10.0	8.2	3.7	0.7	2.2	1.4	0.8	100.0
Ranked #1 (data)	18.1	14.8	13.9	10.5	9.3	8.3	8.4	7.2	4.3	1.1	2.1	1.3	0.9	100.0
Preferred and Ranked #1	16.7	11.7	9.2	9.2	7.8	6.0	7.9	7.0	2.9	0.7	2.2	1.4	0.8	83.3
Preferred and Assigned	10.1	9.2	6.9	7.4	7.1	4.8	7.8	6.8	2.7	0.7	2.2	1.4	0.8	67.9
Ranked #1 and Assigned	10.8	11.1	9.3	8.5	7.9	6.1	10.0	8.0	3.1	0.7	2.2	1.4	0.8	79.8

Table 9: Deferred Acceptance vs Cambridge

	Truthful			Sophisticated		
	All Students	Paid Lunch	Free Lunch	All Students	Paid Lunch	Free Lunch
	Panel A: Deferred Acceptance					
Assigned to First Choice	66.74	56.95	86.12	70.00	60.53	88.74
Assigned to Second Choice	12.93	15.20	8.44	14.82	17.78	8.96
Assigned to Third Choice	5.95	8.52	0.85	4.36	6.15	0.81
Assigned to Fourth Choice	3.58	5.33	0.11	1.07	1.57	0.08
Assigned to Fifth Choice	2.15	3.23	0.02	0.16	0.24	0.01
	Panel B: Cambridge Mechanism					
Assigned to First Choice	78.09	73.65	86.88	74.38	67.44	88.13
Assigned to Second Choice	6.97	7.17	6.58	13.26	15.77	8.27
Assigned to Third Choice	3.21	4.09	1.48	3.24	4.15	1.46
Assigned to Fourth Choice	0	0	0	0.84	1.1	0.32
Assigned to Fifth Choice	0	0	0	0.13	0.18	0.05
	Panel C: Deferred Acceptance vs Cambridge					
Mean Utility DA - Cambridge	-0.02	-0.04	0.01	-0.10	-0.15	0.00
Std. Utility DA - Cambridge	0.41	0.48	0.21	0.29	0.33	0.14
Percent DA > Cambridge	51.49	51.38	51.71	50.5	48.92	53.63

Notes: Panels A and B present percentages of students assigned to true k-th choice. Panel C compares the expected utility difference between Deferred Acceptance and Cambridge Mechanism. Simulations of the Deferred Acceptance mechanism draw other student reports using the estimated utility distribution. All simulations based on the posterior means of the parameters and 1,000 draws.